DOCUMENT RESUME

ED 201 987 CS 006 085

AUTHOR	Johnston, Peter
TITLE	Implications of Basic Research for the Assessment of Reading Comprehension. Technical Report No. 206.
INSTITUTION	Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY	National Inst. of Education (ED), Washington, D.C.
PUB DATE	May 81
CONTRACT	400-76-0116
NOTE	162p.
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	*Evaluation Criteria; Literature Reviews; Measurement Techniques; Models; *Reading Comprehension; *Reading Research; *Research Methodology

ABSTRACT

Emphasizing the work done at the Center for the Study of Reading in Urbana, Illinois, this paper draws together experimental and theoretical information that is relevant to the problem of assessing children's reading comprehension. The first two sections of the paper present and support a model of reading comprehension that integrates some previously opposed perspectives. Having described reading comprehension and the various factors affecting it, the next sections of the paper deal with the methodological problems of gathering useful assessment data, suggest some alternative approaches, and address the question of why researchers might want to assess reading comprehension. The concluding section of the paper contains a summary and implications for research on reading comprehension assessment. (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED201987

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

CENTER FOR THE STUDY OF READING

Technical Report No. 206

IMPLICATIONS OF BASIC RESEARCH FOR THE ASSESSMENT OF READING COMPREHENSION

Peter Johnston
University of Illinois at Urbana-Champaign

May 1981

University of Illinois
at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Massachusetts 02238

The research reported herein was supported in part by the National Institute of Education under Contract No. HEW-NIE-C-400-76-0116.

CS 006 085

EDITORIAL BOARD

Peter Johnston, Chairperson

Roberta Ferrara

Scott Fertig

Nicholas Hastings

Asghar Iran-Nejad

Jill LaZansky

Jim Mosenthal

Ann Myers

Andee Rubin

William Tirre

Paul Wilson

Peter Winograd

Michael Nivens, Editorial Assistant

Acknowledgements

This paper is the result of a decision by the Scientific Advisory Council of the Center for the Study of Reading to produce a paper drawing out the implications of the Center's work for the assessment of reading comprehension. The work was supervised by a subcommittee consisting of David Pearson, George McConkie, Bob Linn, and Andrew Ortony, each of whom gave his time unstintingly. They contributed to both the content and structure of the paper. However, the committee cannot be held responsible for the final product. Thanks are due to Peter Winograd and Rob Tierney for comments on earlier drafts of the paper. Special thanks to George McConkie for his efforts above and beyond the call of duty.

Table of Contents

Page

Abstract 1

DEFINING READING COMPREHENSION 4

The Relationship between Comprehension and Reading: Some Theoretical Issues 5

Issue 1: Comprehension: A Process or a Product? 6
Issue 2: Comprehension: Divisible or Indivisible? 7

An Emerging Perspective on Reading Comprehension 11

Towards a Resolution of the Issues 16

Reasoning Strategies in Reading 17
The Relationship between Reading and Background Knowledge . . . 22

Summary 26

FACTORS WHICH INFLUENCE READING COMPREHENSION AND ITS ASSESSMENT . . 29

The Text 29

Content/Structure Research 31
The Writer-Reader Relationship 36

The Background Knowledge/Text Relationship 41

Quantitative Differences 42
Qualitative Differences 44

The Task 48

Production Abilities 48
Memory and Retrieval 49
Reasoning 50
Motivation 51
Purpose 52
The Social Setting and Interaction 52
Expectation and Perceived Task Demands 53
Test-Wiseness 54

Summary 55

PURPOSES OF ASSESSMENT 55

Administrative 56

Diagnostic 56

Selection and Classification 57

Page

ASSESSMENT METHODOLOGY 58

Constraints upon Assessment 58

Practical Concerns 58
Item Sampling and Generalizability 59
Passage Dependence 62
Item Generation, Selection, and Classification 64
Reliability and Validity 69

The Cognitive Demands/Information Content of Assessment Methods 75

Product Measures 76
Process Measures 85
Metacognitive Measures 90
Summary 95

CONCLUDING SUMMARY AND IMPLICATIONS 96

What Can and Should We Assess? 97

How Might We Assess These Aspects? 101

Question Construction 107

Characteristics and Interpretation of Tests 109

Administrative Assessment 115
Diagnostic Assessment 115
Selection Assessment 117

Reference Notes 119

References 122

Abstract

This paper represents an attempt to draw together experimental and theoretical information which is relevant to the problem of assessing children's reading comprehension. Although the literature review is not exhaustive, it does draw on work from a number of disciplines. The major focus is on the work done at the Center for the Study of Reading.

The biggest contribution of basic research to the assessment of reading comprehension is the increased understanding of the object of the assessment: comprehension. An attempt is made to resolve some of the conflicts which have beset reading comprehension assessment, and to draw out the implications for future study and application.

Implications of Basic Research for the Assessment of Reading Comprehension

In a hall a large group of children are seated at desks working feverishly on optical scan papers. An adult paces up and down the aisles, pausing occasionally, and keeping a constant eye on the clock on the wall.

In a small laboratory, a child is seated in front of a CRT screen which displays segments of text. His head is held tightly in place. The experimenter stands close by and a computer in the background flashes a group of small lights.

In a classroom, a teacher and a child discuss a book which the child has just read. The teacher is asking the child many questions. The noise level in the classroom gradually increases.

What these three scenarios have in common is that they are all concerned with the assessment of reading comprehension. People assess reading comprehension for a wide variety of reasons under a variety of different circumstances, each with a different set of constraints. But assessment has been fraught with problems, a considerable number of which have stemmed from the lack of a coherent theoretical framework, especially one enabling understanding of exactly what reading comprehension is and what it involves. For example, the validity of tests, especially the diagnostic ones, is suspect since there is no complete model with which they can be compared. There is no rational method of test-item generation, and even attempts at item selection have so far been incomplete. Thus, the relationship between questions, text, and the act of reading has been ill-defined and unsystematic. While traditional measurement theory does give information for decisions about which of a set of items discriminates best between individuals, it does not help us select items that best differentiate component skills or that best measure the underlying construct of reading. It certainly does not help us to generate the items in the first place.

Lumsden (1976) claims that we appear to have approached the functional limits of mathematical and statistical sophistication in educational measurement. Tuinman (1979) argues similarly, presenting ample evidence to support the contention and claiming that test constructors have only been saved from total embarrassment by the fact that most tests load heavily on a general verbal factor.

Tuinman has also amassed evidence to show that attempts at forming and validating skill hierarchies have failed miserably. He states that "every teacher knows that it is possible to ask an easy ´higher order´ question and a very difficult ´lower order´ question." (p. 40). The tests bear little relationship to the actual societal reading demands, and since interpretation of tests is not clear from the test development, they are open to misuse and abuse. Recent research may help us in our attempts to solve some of these difficult problems.

This paper is an attempt to draw together experimental and theoretical information which is relevant to the problem of assessing children´s reading comprehension. The emphasis is especially on the work done at the Center for the Study of Reading. The goal is to present the implications of recent research for the assessment of reading comprehension.

One cannot readily cleave the problem of reading comprehension assessment into independent subproblems since the interrelationships between the different aspects are manifold and complex. However, the paper will be divided in the following manner. In order to address the assessment of reading comprehension, there must first be some agreement on how to define the object of our measurement, since the method and criteria of assessment must be governed by one's conception of the dimensions of the object of assessment. An attempt will then be made to delineate the various factors which will influence the outcome of our assessment. Consequently, the first two sections of this paper present and support a model of reading comprehension which integrates some previously opposed perspectives.

Having described reading comprehension and the various factors affecting it, the methodological problems of gathering useful assessment data will be dealt with, and some alternative approaches suggested. A further section will address the question of why one might want to assess reading comprehension. Finally, a summary will be presented and some implications for reading comprehension assessment will be suggested.

DEFINING READING COMPREHENSION

Attempting to define reading comprehension is always a bold move — dissenters are likely to be many. Yet it is a problem which must be addressed, since if we cannot agree on what we are measuring, it seems unlikely that we will obtain any concensus on how to measure it. *It is in* this area of reading comprehension assessment, i.e., understanding the nature of comprehension itself that most progress has been made over the last few years. In order to demonstrate this impact, old and new perspectives on some important issues will be contrasted.

The Relationship between Comprehension and Reading: Some Theoretical Issues

To begin with, reading will be considered as any reader interaction with text. Comprehension is one aspect of reading -- the one which we are concerned with here. Other aspects include: decoding, scanning (e.g., telephone directory use), and vocalizing the print on the page (word calling). This means that assessment of decoding is not here considered as part of assessment of comprehension except insofar as it might be predictive of problems with reading comprehension. This is not to say that decoding is not a necessary prerequisite for reading comprehension, and interactively involved with it. It is both of these.

In attempting to develop a definition of reading comprehension, the first issue to be addressed deals with whether reading comprehension should be viewed as a process or as a product. That is, should it be viewed as the change in (or state of) knowledge which has occurred through reading, or as the process by which the change (or state) comes about. A second issue relates to the nature of the process(es) involved in comprehension. Are there identifiable subsystems operating, or is there one complex indivisible process operating? A conceptualization of reading comprehension and how it helps us toward a resolution of these issues will then be examined.

Issue 1: Comprehension: A Process or a Product?

Most current approaches to the assessment of reading comprehension embody an assumption that reading comprehension is a product of a reader's interaction with a text. This product is stored in the reader's memory and may be examined by convincing the reader to express relevant segments of the stored material. Such an approach is less of a theoretical than a pragmatic attempt at finding out how the interaction has changed the reader's knowledge. The process(es) involved in getting there are given less emphasis than the final product (i.e., the contents of memory). This product position implies that long-term memory plays a large part in comprehension, determining how "successful" the reader was at comprehending, and is typified by standardized tests and free recall measures.

In contrast, Carroll (1971) contends that comprehension is a process which occurs immediately on reception of information and that only short-term memory is involved. Consequently, he states, "As soon as longer time intervals are involved in the testing of comprehension, there is the possibility that we are studying memory processes along with, or in place of, comprehension processes" (p. 6). Carroll's "process" is represented by, for example, eye-movement and reaction-time studies, and miscue analysis.

Royer and Cunningham (1978, p. 36) take issue with both the process and product approaches and contend that ". . . comprehension processes and memory processee are inextricably intertwined. . . . We assume that a comprehended message will be retained in memory better than an uncomprehended message. There is ample evidence to support this assumption (e.g., Bransford & Johnson, 1972; Bransford & Johnson, 1973; Dooling & Lachman, 1971)." Indeed, Bower (1978, p. 212) states "Superior memory seems to be an incidental byproduct of fully understanding a text." This indicates a blend of the concerns of both product and process approaches, an interest in the memorial outcome and the process(es) of attaining it. However, should we construe this blend as being a fact of biology or a flaw in our assessment? That is, are memorial, retrieval, and comprehension processes truly inextricably interwoven, or is it just that we have difficulty assessing them independently?

These three theoretical positions have produced a good deal of friction in the assessment world. The last of the three positions approximates the position which will be presented and justified in this paper.

Issue 2: Comprehension: Divisible or Indivisible?

Reading comprehension has alternately been considered as a wholistic process (Thorndike, 1974) and as a process composed of distinct subprocesses (Davis, 1972; Drahozal & Hanna, 1978). The skills or component approach to reading comprehension is based on the assumption that comprehension can be analyzed into various discrete subprocesses, all of which are necessary for successful performance of mature reading. Further, it is assumed that if a child lacks the skills, then specific remedial training can correct the difficulty. On the other hand, proponents of a "wholistic" view of reading comprehension contend that one cannot break down into components what the mind does. This is a much more pressing, difficult, and deep-seated issue than the process/product issue because it has even more powerful consequences for teaching and remediation.

In the following discussion, "subskills" and "subprocesses" will be used in their broad sense and somewhat interchangeably.

The basic arguments. Many different sets of "necessary and sufficient" subskills of reading have been derived by logical analysis (e.g., Baker & Stein, 1978; Rystrom, 1970; Vernon, 1962) with considerable lack of agreement on the matter. However, the major approach to isolating such skills has been through factor analysis. Davis (1944) presented a rational and statistical analysis of a study claiming eight separate subskills. He reanalyzed his data on several occassions and still claimed small amounts of unique variance accounted for after the first major factor (Davis, 1968, 1972). Reanalysis of Davis′ data by Spearitt (1972) indicated four separate factors: recalling word meanings, drawing inferences, recognizing author′s purpose, attitude, tone, and mood, and following the passage structure. Studies by Bateman, Frandsen, and Dedmon (1964), Drahozal and Hanna (1978), and others have found variety both in type and number of factors, and little consensus on either.

Proponents of the single factor (wholistic) concept have used similar empirical techniques to demonstrate opposite ends (Clark, 1972; Thorndike, 1974; Derrick, Note 1). However, the main proponent of the wholistic approach has been Thorndike (1974, p. 57) who claims "The barrier . . . [is] . . . not primarily a deficit of one or more specific and readily teachable reading skills but is a reflection of generally meager intellectual processes." Thorndike uses a variety of data to support his contention that reading is in fact reasoning and cannot be decomposed into separate elements.

Problems with the subskills/wholistic research. There have been a number of other problems associated with the research used to substantiate the subskills and wholistic positions, aside from the common arguments (stated above) over type of communality chosen and the types of rotation used in the factor analyses. These problems include the following:

1. The test passages appear to have been arbitrarily chosen.
2. Analyses have been across different age groups. However, different skills may be important at different ages, and different items may be interpreted in developmentally different ways (e.g., Stein & Glenn, 1978).
3. Items have been selected on the basis of their capacity to discriminate between individuals rather than their capacity to differentiate separate skills.
4. The word knowledge factor probably represents global world knowledge.
5. Factor analysis depends on item intercorrelations and these depend on variance of the traits in the sample. If there is one trait and small variance, one can get a number of unstable factors. But if the traits measured are correlated, and subject variances on the traits are large, then the first factor smothers the others. This is almost bound to occur given the facts expressed in points 3 and 4 above. Also, if the items are a hierarchical set (as is

indicated by Davis, 1972, and even showing up in Thorndike's, 1974 results), subjects able to answer higher level questions should be able to answer lower-level questions. Thus, if there is good variation in the subjects' abilities, then inter item correlations will be high and, again, the first factor will dominate (Andrich & Godfrey, 1978-79).

6. The search has been directed at finding independent skills and there seems to be no reason to suppose that different reading comprehension skills should be uncorrelated. Indeed, interfacilitation of subskills has been indicated by Guthrie (1973).

These problems are not restricted to the factor analytic approaches to the location of subskills. For example, Andrich and Godfrey (1978-79), used a single-logistic model rather than a factor analytic model. With Davis' data they located four hierarchical subskills which differ from the other analyses but which still lack logical or intuitive appeal.

No doubt some of the motivation to find some separable subskills has arisen from the fact that without them, diagnostic assessment would be essentially meaningless, since there would be nothing to diagnose separately. Part of the problem here is that without a theory that can guide the search for identifiable subskills, it has been difficult to get a feeling for the "level" at which to look for them. For example, "main idea finding" is a popular conception of a subskill, but so is "word knowledge." However, if we look more carefully at the nature of the comprehension task, we may find (as is argued below) that we have been looking for the subskills in the wrong places.

It is clear that it would be useful to know of a set of remediable subskills which comprise reading comprehension. This would provide a framework for effective assessment and remediation of reading comprehension difficulties. If we are to locate such a set we must conduct a search that is driven by theory rather than by solely pragmatic concerns. Thus we must begin to elaborate a coherent theoretical model of reading comprehension.

An Emerging Perspective on Reading Comprehension

Recently there have been some basic changes in our conception of reading comprehension. However, these have not yet been reflected in our assessment procedures. In 1971, Carroll posed some questions to researchers. He asked

> . . . whether it is possible in fact to distinguish "pure" comprehension of language texts from processes of inference, deduction, and problem solving that often accompany the reception of language. An empirical question would be to see whether it would be possible to decrease the correlation of comprehension ability tests with intelligence tests by eliminating or reducing those elements of comprehension tests that call for inferential processes that go beyond sheer comprehension. . . . Research is needed to see to what extent it is possible to reduce their dependence on memory. (p. 3)

In the past eight years, much of the appropriate research has been done, and the result has been that the original questions are no longer relevant. For example, Pearson and Johnson (1978, p. 24) contend that

> Comprehension is building bridges between the new and the known. . . . Comprehension is active not passive; that is, the reader cannot help but interpret and alter what he reads in accordance with prior knowledge about the topic under discussion. Comprehension is not simply a matter of recording and reporting verbatim what has been read. Comprehension involves a great deal of inference making.

The contrast between these quotes by Carroll and Pearson and Johnson indicates the recent changes in our conceptions of what comprehension is. While there is still some willingness to separate inferencing and problem solving, they are no longer generally considered to be removable from the comprehension process, but rather are considered an integral part of it, like the apple in apple pie. One must even infer the intended meanings of single words in context (especially polysemous words like "run"). Indeed, the status of inference has shifted over the last few years from a single process, almost an optional extra, to a selection of fairly well-differentiated types of inference upon which virtually all comprehension is predicated.

We do not consider that someone has comprehended something if they can only give a rote recall of the elements. We only consider that a reader has comprehended a text when he has established logical connections among the ideas in the text and can express these in an alternate form. In this way, inferences are critical acts of comprehension, since they allow us to make various words meaningful, join propositions and sentences together, and fill in the missing chunks of information.

Pace (Note 2) has pointed out that in familiar story situations there is a difference between "active" and "passive" inferencing, passive inferencing being a matter of merely "recognizing" the appropriate schema. Thus it is clear that the common "inference question" is no longer a single question type. Indeed, many "literal" questions may involve inferencing, especially at the lexical level. We must consider what type of cognitive process was involved, based on our growing understanding of these processes.

Trabasso (1980) has pointed out that the functions which these inferences perform are fourfold:

1. resolution of lexical ambiguity;
2. resolution of pronominal and nominal references;
3. establishing context for the sentence;
4. establishing larger framework for interpretation -- a model base for top-down processing.

These functions are here ordered from bottom-up to top-down in terms of the processes which they involve. Top-down processing requires a previously formed knowledge structure which already contains the major relationships. This is conceived of as a "slot-filling" activity. That is, the appropriate knowledge structures already exist, only the specifics need to be "filled in." In the case of an absence of such a structure (or the failure to use it) the reader must rely more on bottom-up processing, requiring more use of lexical and linguistic knowledge and word recognition to build meaning sentence by sentence, and thence proceed to build a higher order structure.

Warren, Nicholas, and Trabasso (1979) have a more detailed analysis of inferencing in terms of what the inferences do. They can be slot-filling if they generate extra information which was not actually in the text. For example, if the text gives "James flew to New York," then one infers that James probably went to the airport, entered the plane, etc., first. Learning that James went in a taxi is merely providing expected information, since there is already a "slot" for such information. If the information is not provided, we provide it ourselves. Alternatively, inferences can be text-connecting if they establish connective relationships between elements of text. There are a number of subclasses of inferences, independent of this rather text-based classification (slot-filling versus text-connecting). These are:

1. Logical relationships:
    a. Motivational. For example, the text says "Bill had not eaten in two days." One is likely to infer from that a certain motivation on Bill's part to find some food.
    b. Enablement. One can infer without difficulty that wealth enables purchasing of things.
    c. Psychological cause. One could readily infer, given the appropriate background that one person's hatred could have been the cause of someone else's death.

d. Physical cause. One can infer that an ice-laden road caused a car to skid.

2. Informational relationship:

   a. Spatial and temporal. Given that A occurred before B, it can be inferred that B occurred after A.

   b. Pronominal and lexical. Knowing who "he" refers to in the sentence and what "run" means.

3. Evaluative. Inferences which are based on moral and social judgement, such as John beats his wife; therefore, he is not a nice person.

Such inferences are the very essence of reading comprehension and the more the reader makes them, generally, the more he has comprehended. For example, Omanson, Warren, and Trabasso (1978) found that by giving children clear information on a protagonist's goals, thus allowing more inferences to be made, the children showed greater comprehension of the stories.

However, this is not to say that the reader is to generate all possible inferences. This would lead to a complete loss of the author's message. The reader, rather, has a system for directing inferencing. This system seems to utilize a concept of "good form" such that what an author leaves out of a text, and what a reader is allowed to add, are governed by some implicit canonical forms and Gricean-type rules (Adams & Bruce, 1980).

It is likely that inferencing is initially fairly text-based but that good readers quite quickly shift to a more model-based generation of inferences. That is, since the possible inferences are often many, the inferencing processes must be somewhat selective. It seems that we try to get a jump ahead of the text by inferring where it is leading us, building a mental model of what we think it is about. Our model produces questions which need to be answered ("slots to be filled") and as we answer these there are fewer alternatives for other questions, and new questions arise. If we find an answer or another piece of information which does not fit our model, we have to examine the assumptions which we have made in constructing our model.

These processes are conceived of as the reader's active interaction with a specific text in the context of his background knowledge and a given task, and in a given social setting. This concept will be elaborated throughout the paper.

In a nutshell then, reading comprehension is viewed as the process of using one's own prior knowledge and the writer's cues to infer the author's intended meaning. This involves a considerable amount of inferencing at all levels as one builds a model of the meaning of the text. If prior knowledge is strong then a detailed model may be rapidly constructed which reduces the reading to slot-filling and verifying, and inferences are merely default values in the model (elaborated on p. 18 of this paper).

Towards a Resolution of the Issues

The theoretical perspective presented above can be substantiated and elaborated further using evidence from several recent theoretical and experimental sources. Through this elaboration the discussion is intended to demonstrate that earlier research and conceptions of reading comprehension were restrictive. Those few researchers who considered reading comprehension in the context of general cognitive abilities were hampered by the lack of adequate research on cognitive skills.

Considering reading as a specific application of more general cognitive processing skills and strategies enables us to propose some resolutions to the theoretical issues presented earlier in this paper. For example, the conflict between Thorndike's (1974) claim that reading is, in fact, reasoning, and therefore not a series of subskills, and the subskills approach can be reduced by recognizing that reasoning is not an indivisible ability. Thus, while reading may well be reasoning, that does not preclude it from having subskills.

Reasoning Strategies in Reading

Several lines of research indicate that "reasoning" is not a unitary ability. Seigler (in press), in studying children's problem-solving behavior, looked at the strategies which children were able to use to solve various problems. He was able to analyze the strategies used and diagnose and correct problems with them. Recent work by Sternberg (1977) also indicates that one can define a series of strategies for answering various I.Q. test questions. Furthermore, training subjects in the use of these strategies leads to great improvements in performance. Similar results have been noted by Pellegrino and Glaser (in press).

That reading involves reasoning is apparent from a number of studies in artificial intelligence (e.g., Schank & Abelson, 1977; Wilensky, Note 3) and reading comprehension (e.g., Collins, Brown, & Larkin, 1977; Collins &

Smith, 1980). The direction of research is now toward making explicit the reasoning strategies involved in reading comprehension.

There are two general "levels" of strategy which can be described. The first set of strategies is involved with helping the reader construct a model of the meaning of the text from the information available in the text and in the reader's head. The second set is that by which readers monitor their progress toward understanding the text, detecting lapses in comprehension, and initiating strategies to rectify the difficulties.

Strategies involving cue use and model building. On the basic level, reading comprehension involves the use of textual cues to create a model of the presumed meaning of the text. There is a wide variety of different cues and information available to readers to assist them in their attempts to construct the meaning to which the symbols on the page point. Information is available from the reader's head (background knowledge), sometimes from the social setting, and from the page. The information takes certain systematic forms, and there are many different relationships between the information sources.

Meyer's (1975) and Clements' (1976) works indicate that good readers have an awareness of different types of macro-signals within the text. Macro-signals are cue systems which give information about the organization of the information on the page. Brown and Day's (Note 4) work shows that readers can be made aware of other structural aspects of text such as repetition and elaboration. An analysis of several further important cue systems has also been presented by Collins and Smith (1980). (See the later section on Text in this paper.) Readers can use such sources of information to construct their model of the text meaning and to evaluate it. Collins, Brown, and Larkin (1977) present some of the strategies which readers use to do this:

1. Rebinding. If the interpretation of the last piece read causes a conflict in the model, then keep all of the model to date but find a new interpretation for the last bit.
2. Question default interpretation. The type of question here is "Is X a standard X with standard properties?"
3. Questioning a direct or indirect conflict. For example, a chain of inferences leads one to a conflict with another part of the model, which forces one to return and question an earlier inference in the chain.
4. Near or distant shift of focus. The reader reaches a problem which he cannot solve, so tries to answer a different question.
5. Case analysis and most likely case assignment. Sometimes readers fill in with guesses and then check to see whether the consequent constraints allow convergent solution.

In using the above strategies to evaluate their models, readers also have certain conditions for acceptance or rejection of the total model. They check (generally unconsciously):

1. the plausibility of the assumptions and consequences of the model;
2. the completeness of the model;
3. the interconnectedness of the model;

4. the match of the model with the text.

That is to say, in testing the model, readers seem to check its predictive power, whereas it is constructed by checking on the fit with prior data.

Kintch and van Dijk (1978) presented a theoretical conception of how information is processed while reading and used variation in some of the model's parameters to account for individual differences in recall. The model includes a set of summarization strategies. Brown and Day (1980) analyzed the think-aloud protocols of proficient summarizers and validated a similar set of strategies. These researchers then taught children and adults the strategies for finding the "gist" of a passage and found that the subjects' ability to summarize improved immensely. They essentially found that people look for certain information and then initiate a processing strategy based on what they find. Some examples are, "Is there a topic sentence for this paragraph? No? Then construct one. Is this information repeated? Yes? Then delete it."

Brown and Day's work indicates that by instructing readers on the appropriate information checklist and rules to use, readers' summaries can be improved. While these strategies are initially used consciously, Brown and Day found that experts at summarizing used them without being aware of doing so. That we are dealing here with the use of separate strategies (cognitive skills?) is given added weight by the fact that their analyses show that some children consistently used some strategies but not others, and that some of these strategies appeared to be more difficult than others. Thus it seems that what had been one of the few intuitively appealling "subskills" (i.e., "finding the main point"), may actually break down into several lower-level strategies which are separable and teachable. More will be said about these issues when we consider the task demands of different types of reading.

Given this conceptualization of reading comprehension, it is eminently clear why there are such high correlations between tests of reading comprehension and tests of reasoning. Reading comprehension does have a large reasoning component.

Strategies involving comprehension monitoring. There has recently been a great deal of research into comprehension monitoring or "metacomprehension" (e.g., Baker, 1979; Brown, 1978; Markman, 1979; Winograd & Johnston, 1980).

Comprehension failure can occur at the word, sentence, intersentence, or discourse level. At each level there is the problem of being unable to locate a reasonable interpretation, or of having several possible interpretations. At the discourse level, one can also fail to "get the point" or fail to understand some section in relation to some other section, and so on. What is required in comprehension monitoring is the awareness of the "triggering conditions," or symptoms of comprehension failure, and an awareness of the strategies which can be used to remedy the situation. Collins and Smith (1980), in presenting some of these strategies, also point out that each repair strategy carries a cost, since it either diverts the reader's attention from the main thread, or it puts the reader at risk of failing to understand a larger segment of the text. Some of these strategies are:

1. ignore and read on;
2. suspend judgement;
3. form a tentative hypothesis;
4. reread the current sentences;
5. reread the previous context;
6. go to an expert source.

The use of these strategies will obviously depend upon the purpose for which the reader is reading. For example, a reader reading for "gist" will be more inclined to "ignore and read on" than to "reread the current sentence."

The Relationship between Reading and Background Knowledge

Invariably, factor-analytic studies of reading comprehension have found a word knowledge factor on which comprehension tests load highly. In studies of readability, too, any index of vocabulary difficulty accounts for about 80% of the predicted variance (Coleman, 1971). Models of reading comprehension must account for these facts also.

Anderson and Freebody (1979) have examined the three competing hypotheses which attempt to explain this finding: instrumentalist, aptitude, and background knowledge. The instrumentalist position is that knowing words allows text comprehension and not knowing them means that one cannot adequately proceed through the text. The aptitude hypothesis considers vocabulary knowledge as just another index of verbal I.Q. which is the real factor accounting for comprehension (a wholistic type of approach). The knowledge hypothesis suggests that vocabulary knowledge is a secondary index of the extent of background conceptual frameworks (schemata). While data are not available to allow us to choose among these three hypotheses, the knowledge hypothesis is most consistent with the theoretical framework developed thus far. The more background knowledge the reader possesses, the more likely it is that the reader knows the relevant words, and the more likely he or she will be able to make appropriate inferences while reading, and to build appropriate models of meaning. Indeed, Spiro (1980) has proposed a series of "subskills" which derive from a schema-theoretic notion of reading comprehension. This includes consideration of the use of background knowledge and the integration with background knowledge of what is derived from the text.

Spiro's analysis arrives at the following possibly distinct areas:

1. Schema availability--i.e., presence or absence of background knowledge;
2. Schema selection--i.e., decisions on the selection of an appropriate schema to fit the data to;
3. Schema maintenance--i.e., maintaining activation of the selected schemata while proceding through the text;
4. Schema instantiation and refinement--i.e., using the appropriate schema to organize data and refine the model which one is building to fit the data;
5. Schema combination--i.e., integration of different sets of knowledge;
6. Non-analytic aspects of schema based processing--i.e., the subjective "feelings" associated with processing.

This approach is not in conflict with the notions presented above; rather, it complements them. For example, the Kintch and van Dijk (1978) model requires schematic knowledge in order to operate effectively. Such knowledge aids in reducing the short-term memory load in the model since information can be transferred more rapidly from short-term memory to long-term memory and information can be stored in larger chunks (Chi, in press).

Spiro's model contains concepts of schema availability and schema selection or access. Schema availability refers to the problem of lack of the requisite background information about a specific area, whereas the problem of schema selection refers to the failure to use available background knowledge when it is required.

Other "subskills" Spiro presents represent the strategic use of available information. For example, schema selection has been looked upon as a trainable skill by Winograd and Johnston (1980). The evidence suggests that provided the schema is present in the first place, selection is a learnable skill. Some children do not relate what they read to their background knowledge but can be taught to do so. Spiro (1980) also suggests that updating old schemata rather than continually attempting to construct new ones is a skill which a child does not necessarily bring to the reading situation. Schema maintenance is important too, since often one must maintain certain referents in working memory in order to maintain coherence.

Bransford (1979) has provided a good illustration of the effects of teaching children strategies for schema use in reading. Children who used these strategies exhibited excellent recall in contrast to those who did not. Furthermore, children taught to use the strategies were able to improve their recall considerably.

The upshot of this work on prior knowledge via schema theory is that the large correlations between reading comprehension tests and vocabulary tests are quite predictable. If vocabulary is simply indicative of background knowledge, then limited vocabulary would indicate a limited ability to produce the appropriate inferences and build the models of meaning which allow more efficient top-down reading. That is, the reasoning component of reading would be restricted by the lack of available data to solve the puzzles in the text. On the other hand, more reasoning skill would be required to make the text coherent.

Brown, Campione, and Day (1980) have distinguished between these three important types of knowledge; strategic, content, and metacognitive. Strategic knowledge refers to the repertoire of rules, procedures, tricks, routines, etc. for making learning a more efficient activity (Brown, 1975). Content or factual knowledge refers to information that learners have concerning the subject under consideration and their general knowledge of the world (Anderson, 1977; Brown, 1975; Chi, in press). Metacognitive knowledge refers to the information that learners have concerning the state of their own knowledge base and the task demands they are facing (Baker & Brown, 1980; Brown, 1975; Flavell & Wellman, 1977).

This distinction links the present section on prior knowledge with the previous section on reasoning subskills, since it is clear that a reader must have knowledge of appropriate strategies but may or may not make use of the knowledge. It also introduces a further component, that of knowledge about the demands of the task to be engaged in. That this is related to reading failure has been shown by Myers and Paris (1978) and Canney and Winograd (1979). For example, children may be unaware that the purpose of reading is to gain information from text, or that it is necessary to use what they already know in order to understand the text.

Summary

Our concept of reading comprehension has changed quite radically over the last few years, along with our methods of studying it. It is hypothesized that knowledge is stored in schematic structures (Anderson, Spiro, & Anderson, 1977; Rumelhart & Ortony, 1977), and *comprehension* is the process(es) involved in forming, elaborating, modifying or integrating these knowledge structures (Rumelhart, 1977). Degree of comprehension of a text can be considered in terms of the creation of, or degree of appropriate modification, elaboration, and integration of, the relevant knowledge structures. That is, it refers to the extent to which the information represented by the discourse is represented in these cognitive structures. But more than that, it refers to the extent of the interrelationships between and within the structures. (This is not restricted to factual information but includes emotive information also.)

Reading comprehension is considered to be a complex behavior which involves conscious and unconscious use of various strategies, including problem-solving strategies, to build a model of the meaning which the writer is assumed to have intended. The model is constructed using schematic knowledge structures and the various cue systems which the writer has given (e.g., words, syntax, macrostructures, social information) to generate hypotheses which are tested using various logical and pragmatic strategies. Most of this model must be inferred, since text can never be fully explicit and, in general, very little of it is explicit because even the appropriate intensional and extensional meanings of words must be inferred from their context.

In addition to the reasoning processes, good readers monitor the progress of their comprehension, and use fix-up strategies when necessary. This also requires them to decide on the basis of their purpose for reading when to remove their processing from "automatic pilot," take conscious control, and instigate the appropriate alternative strategies. This is a very important aspect of reading comprehension.

The position largely resolves the issue of viewing reading as reasoning versus as a series of subskills. However, by demonstrating that, in fact, reasoning itself has subskills, the two positions are not incompatible. Reading can be reasoning and have subskills at the same time.

The position does not, however, fully resolve the process versus product debate, but rather places it in a different perspective, reducing its relevance somewhat. The problem becomes one of largely being stuck with product measures but being more interested in the processes (particularly for the purposes of diagnosis), since these are what we can affect through instruction. It seems that the two approaches to assessment (process and product) should not be dichotomized but rather should be considered complementary approaches to the same problem. This view is based on the assertion that elements of both process and product exist in many forms of assessment and that reading comprehension will be impeded by failures of processing, storage, or retrieval of information.

Since it is the processing aspects that educators can, in principle, alter through instruction (little can yet be done about storage per se), these become important objects of assessment. It is thus unfortunate that the most readily available forms of assessment are of the product type. However, it is clear that the processes of reading comprehension _must_ utilize or act upon the already stored information, so we are forced to study process and product in their interaction. Indeed, this paper supports the position taken by Royer and Cunningham (1978) and holds that studying comprehension without studying memory is somewhat irrelevant. It is rather like contemplating the sound of the proverbial tree falling in the primeval forest. It is philosophically interesting but practically and psychologically of minimal interest.

The theoretical issues presented in this first section, provide a base from which we can develop our understanding of what we are measuring and thence what such a measurement will mean. While this discussion is far from complete, the following sections will expand further some of the issues raised above. The next section is devoted to an examination of the factors which affect reading comprehension and its assessment.

FACTORS WHICH INFLUENCE READING COMPREHENSION AND ITS ASSESSMENT

Assessment of reading comprehension requires interpretation of an individual's performance of some task which is based on information from a given text within a given context. Thus, performance on the test will depend on characteristics of the text, the nature of the task, and the context, as well as the person's reading abilities and prior knowledge. Consequently, it is necessary to understand the influence of these associated factors in order to make a meaningful interpretation of an individual's performance on assessment devices. This section of the paper will deal with the various factors which influence reading comprehension and its assessment. These factors turn out to be excellent places for teachers and others to look to find explanations of children's real or apparent reading comprehension failures. The topics dealt with are:

1. the text in terms of its content, structure, and language;
2. the appropriateness of the text to the student's prior knowledge;
3. the sources of answers to questions;
4. the task demands of the assessment procedures.

The Text

If one has a selection of samples of different texts, it seems likely that the level of difficulty for a given reader will vary considerably across the texts. There are many possible causes for such variability in difficulty, and each potential cause is likely to differ in potency from reader to reader. A fair amount of research has been directed at isolating these variables.

The research has largely involved analyzing text to determine its readability. Though the work is not yet able to algorithmically improve prediction of readability, it can indicate to an extent why one piece is likely to be more difficult than another. However, some of this work is still rooted more in linguistics than in psycholinguistics, and is thus less concerned with the psychological reasons for reading difficulty. Much of this work is also still pragmatically based rather than theoretically based. Text difficulty is currently defined in terms of either normative difficulty or readability formulae (regression prediction of normative difficulty), the latter being based on secondary criteria--pragmatic predictors such as sentence length or word frequency. These measures do not seem able to adequately distinguish between well-written, challenging texts and badly constructed texts which make interpretation difficult.

In order to select passages for assessment procedures, we need to know the characteristics of the passages--what makes one different from another. This is especially so if we wish to test children with a range of ability. Alas, the state of theoretical development in psychology and linguistics cannot support a definitive analysis of text at this point. This is, in fact, the objective of much current research, which is proceding along two lines:

1. analysis of text in terms of its content and structure and the relationships between the two;
2. analysis of the writer/reader relationship, looking at text as a communicative device.

The import of these two approaches will be considered next.

Content/Structure Research

The characteristics of texts have been studied from the perspective of text as text (somewhat independent of author). One such aspect of text which affects what a reader recalls is the sheer quantity of information, as measured by passage length (Newsome & Gaite, 1971), density of information (Goetz, Anderson, & Schallert, 1979; Kintsch & Keenan, 1973), and density of new information (Aiken, Thomas, & Shennum, 1975). Kintsch, Kozminsky, Streby, McKoon, and Keenan (1975) have also shown that the density of arguments in propositions slows reading and reduces recall. The more text content is concrete and imageable (Paivio, 1971), and interesting (Johnson, 1974), the more memorable it is also. Bower (1978) also shows that text can induce a reader to identify with a particular character, thus influencing recall.

The extent to which information is represented in the text (so that the reader does not have to infer it) has an effect on recall. McConkie (1978, p. 17) summarizes this:

> In general, textual manipulations which reduce information useful to the reader in building a coherent representation of the content can be expected to reduce comprehension of the passage. In some cases, the reader's prior knowledge will compensate for loss of textual information or relations may be identified on a problem solving basis requiring greater reading time.

This clearly indicates the importance of being able to specify, as far as possible, exactly what is and is not stated in the text both explicitly and implicitly, especially if one intends to measure in what way the reader's knowledge changed as a result of his interaction with the text.

Specifying what must be infarred is especially important for developing test items and understanding performance on them, particularly the specific use of background knowledge. For example, there are developmental differences in children's contextual inference-making skills. Recall accuracy is highly correlated with these inferences, and increases with age (Paris & Upton, 1976).

Recently there has been considerable interest in analysis of narrative text. The common elements in different narrative texts has given rise to what have been referred to as "story grammars." These posit the existence of conventionalized macrostrucures derived from background knowledge of texts and the way things work in the world. That even young children are aware of such structure was shown by Brown and Murphy (1975). These experimenters found that 4-year-olds recalled logical ordering of pictures better than random orderings. They also found that if the random orderings had a logical story built around them, recall improved. Thus it seems that even young children can use their knowledge of logical relations between things to help understand and recall information. A considerable amount of information has accumulated on this form, some of which indicates differences between good and poor readers, and developmental differences.

Knowledge of and use of these conventional macrostructures appears to facilitate text comprehension. They seem to have their effect by allowing the reader to rapidly construct from the text a model of a possible world of situations which accounts for relationships within the text. These models have been described variously as frames (Minsky, 1975) or schemata (Rumelhart & Ortony, 1977) and are units of conventional knowledge which provide a basis for the organization of mutual expectations and interactions. These have given rise to the notion of "well-formedness" of text.

While young children possess such structures, they seem less able to use them to organize information. Brown and French (1976) found that preschoolers had more difficulty than older children in remembering sentences when the order in which the sentences were presented differed from the natural order of the events. Stein and Glenn (1978) showed that 5-year-olds are good at the recall of well-formed stories, and Stein and Glenn (Note 5) showed that children deal with systematic disruptions in the well-formedness of the stories in developmentally different ways.

There are cues to these organizational structures in text, for example, topic sentences, macro-connectors such as "but," "however," and "because," and discontinuities in time, location, actors, and content. Provided that a reader is aware of these cues and knows what to do with them, they can assist him in his cognitive modelling of the text. Stein and Nezworski (1978) showed that fifth graders could recall structural disruptions at least as well as normal orderings if explicit markers were used. However, these markers were only partially helpful for first graders. It has also been shown (Marshall & Glock, 1978-79) that, in the case of expository text, some readers are less bound to be influenced by the text structure than are others. More proficient adult readers are able to read text for their own purposes and can override such built-in emphases. This finding has implication for what is considered assessible in reading comprehension. It would seem that perhaps one could assess whether, in fact, a particular reader is sensitive to such structures and whether he is able to override them for his own purposes.

A considerable amount is now known about the characteristics of narrative text, and some work has been done on the structure of expository text. In expository text it is more difficult to describe macrostructures because the relationships between text segments are more diverse than those in the narrative. This is largely a reflection of the greater variety of potential content matter. However, work by Clements (1976), Grimes (1975), and Meyer (1975) has shown that such rhetorical structures do exist in expository text and are used by people in their efforts to construct meaning.

These recent developments have given us greater direction in the area of text analysis. Story grammars have shown the kinds of inferences that people make in certain content/structure situations. The structural expectations which the reader develops are to a considerable extent dependent upon the text topic because of, in the case of narratives, the reader's stored knowledge of the social situations and the normal description of them.

It is probable that readers attempt to form a causal chain as a central core of a passage (Schank, 1975) and consequently, attributive information, which is normally not part of the causal chain, is not as well remembered as causal or action information, which is normally part of the chain. If this more trivial information is what is tested in reading comprehension tests (Tuinman, 1979) then it is important that we know what problems readers have in retrieving it.

On the other hand, if causal information is not stated in the text, it is normally inferred and, because it is stored as part of the causal chain, is likely to be related or identified as having been stated explicitly. But because children are not always able to use such knowledge for top-down processing it is likely that the structure of their recall will differ from the canonical forms.

Some further interesting relationships between the structure and content and and comprehension assessment, were found in a study by Freebody (1980). While he obtained data to show that prior knowledge and word difficulty independently influence reading comprehension, he also found that the order in which subjects read passages affected their comprehension of them. Particularly, he found that variations in cohesion were only effective when they occurred in texts read at the beginning or end of a series of three passages. Evidence of primacy and recency effects were also noted in subjects' performance, independent of the importance of the information. Clements (1976) obtained a similar effect also. This may be important information for the interpretation of current assessment devices which present the reader with a variety of brief passages.

There is little doubt, then, that variations in the structure of text are often related to content and that these variations have distinctive effects on different readers. Readers' failure to effectively deal with these structures and disruptions in them may be diagnostic of broader problems which they may face in their reading.

The Writer-Reader Relationship

The interaction between author and reader via the text, like all social interactions, is based on "good form," and in that way is consistent across texts, though the specific form may vary among different text types.

Beyond the work on narrative and rhetorical structures, work is being conducted by Cohen and others (Adams, Bruce, Cohen, Collins, Gentner, Rubin, Smith, Starr, Starr, & Steinberg, Note 6) on the characteristics of instructional forms, and by Bruce (in press) on the broader, social aspects of text. Both are examining different aspects of text as they relate to the writer-reader interactions.

In this approach the text is considered in its role as a communicative device. The author has a knowledge base, some of which (the passage base) he wishes to communicate to the reader. This passage base is frequently nonlinear in structure, yet the final text must be linear. For example, while the actual events being described originally occurred at the same time, they must be described consecutively. The author's task is to organize this message in light of what he perceives to be the characteristics of the reader, in terms of presumed background knowledge, possible goals, and so forth, to produce a text from which the reader will be likely to derive his intended meaning.

Cohen is attempting to specify the kinds of assumptions which both reader and writer make in their attempts to provide and gather information. He considers the expert-novice instructional interaction to be a special case of social interaction, and by varying the modality of communication, he is able to infer the assumptions which reader and writer make about one another. Reader and writer each attempt to estimate the other's goals and purposes.

Bruce (in press) has used a similar conception of reading as a communicative social interaction. This conception of the reading allows us to consider certain dimensions of the text: participants, meaning, time, location, and physical text. It is possible to have levels of interaction between author and reader, between character and character, and between levels of meaning, and interactions between each of these.

While considerable work has been done on the development of certain of the structures which occur in texts, less work has been done on the problem of constructing a broad system for text classification. Developing such a system is an essential project because, recognizing that different types of text tend to represent different tasks and different problems for the reader, we need to draw test passages from different domains of text and need to be able to specify the characteristics of the domains.

Bruce's work represents a good beginning in the development of a classification system. The work classifies narrative text with respect to a psychologically meaningful system based upon an awareness of social interactions. It also gives a key to what a reader needs to be aware of in order to attain various levels of interpretation of complex narratives. At the same time, Bruce's work indicates another cue system of which the accomplished reader should be aware. One might wish to assess the awareness and use of this system also. The general idea is that writers use the tools available to them to create cues which they feel will best help the reader generate the meaning which the writers intend to convey. In turn, the onus is also upon the reader to use these cues to infer (construct a model of) what the author intended. In a sense, accuracy becomes somewhat relative under these circumstances, and to an extent, one must tolerate a "band of interpretation reflecting varying degrees of reader-based and text-based processing" (Tierney & LaZansky, 1980). Accuracy is in this way constrained by the reader's purpose for reading, his background knowledge, and the extent to which he was able to infer the author's intentions.

This author-reader interaction has also been conceived of as an implicit allowability contractual agreement between author and reader (Tierney & LaZansky, 1980). The agreement is similar to that suggested by Grice (1975) for oral communication, in which the speaker should be informative, sincere, relevant, and lucid. The written situation is somewhat different in that the reader cannot inform the writer of his progress in constructing meaning. However, in a sense this merely makes the writer's task more difficult, since he must make assumptions about the reader's knowledge and perspective in advance.

Kantor (1978) adds to this concept what he terms "considerateness." This describes the extent to which the author has lived up to his side of the agreement, in terms of his use of grammaticality and "flow" of the discourse. That is, the author should generally make meaning construction as simple as possible for the reader.

Green (1979) has presented convincing evidence that newspapers represent a conflict of perceived intentions and goals. She argues that the journalists' beliefs about style, content, and format act together to obscure the structure of the story and its internal relationships, thus effectively blocking the journalists' own goals. This approach considers writing-reading as one mode of communication, which in turn is one type of social interaction. Analysis thus requires us to consider the goals and beliefs of the interactants, and what the different text forms are intended to communicate (Bruce, in press).

Both reader and writer have a complex task. Meaning is constructed rather than transferred and both writer and reader play an active part in constructing meaning using the belief that small elements invoke larger, mutually understood schemata. Because the writer makes certain consistent assumptions about the reader and uses his own language and knowledge in consistent ways, familiarity with an author's style may aid reading comprehension. Children are sensitive to such stylistic differences at an early age (Green & Laff, 1980).

In summary, then, we must recognize the relationships between structure and content of text and the relationship between these as they exist on the printed page and as they exist in the head of the reader. The writer endeavors to present to the reader such cues to his intended meaning as he deems appropriate. These he structures in such a way as to best represent what he wishes to convey to the reader. On the other hand, the reader may read for a purpose other than that for which the text was written and impose upon the text his own structure. This is an essential skill for a reader to have. To choose to accept or ignore the author's structure, however, must be an active choice under given conditions and goals.

In terms of the assessment problem, the findings mentioned so far in this section of the paper suggest that if we can clearly manipulate different aspects of text structure, and the explicitness of cues in the text, we may be able to examine which cues a given reader is or is not using, and which structures are causing him problems. Readers could then be alerted to such cue systems or structures. Thus, any proposed text classification system should classify along psychologically meaningful dimensions.

The fact that there is great variation within individuals across texts, and that much of this variation is as yet unexplained, indicates that observation of readers' performances on different types of texts should provide much potentially diagnostic information which we do not yet have access to. Hopefully we are looking in the right places.

The Background Knowledge/Text Relationship

The social, linguistic, and cultural aspects of the environment in which one grows up have their effects on test performance in numerous ways, some of which have only recently been uncovered. For example, children's understanding of a question may differ from that of adults. The manner in which they represent the problem determines their understanding of it and their willingness and ability to answer it. A primary direction of study in this area has been towards examination of the effects of cultural differences.

Schema theory (Adams & Collins, 1977; Anderson, 1977; Rumelhart & Ortony, 1977) has highlighted the different effects of background knowledge in reading and stressed the need for culturally appropriate instructional reading materials both in terms of linguistic content and background knowledge. It may be that cultural matching of early reading materials (and appropriate assessment) may help, but sooner or later evaluation in the "real world" will occur in the standard language. This presents something of a bind in that one could conclude that assessment would be more equitable if it were carried out with text matched to the learner's linguistic background. However, the standard for evaluation may not then be meaningful with respect to real-world tasks. On the other hand, evaluation which included both standard and dialectal American texts could perhaps be used to assess whether or not a linguistic mismatch was causing a reading problem.

A student might perform badly on a reading comprehension test simply because of a mismatch between his background knowledge and the nature of the text. Indeed, a study by Hall, Reder, and Cole (1975) indicates the likelihood of such an outcome by pointing out differences in free and probed recall of young Black children orally presented with text in either Black English Vernacular or Standard English. Such a mismatch can be quantitative or qualitative in nature.

Quantitative Differences

A quantitative mismatch refers to a lack of relevant world knowledge, for example, when a rural child reads a passage about city metro systems or a child who knows nothing about war reads a passage about the Civil War. Indeed, Johnston and Pearson (Note 7) demonstrated that a brief measure of background knowledge can predict comprehension performance on a given text better than standardized reading test scores can.

The source of difficulty in such a case is clear from the description of reading comprehension presented in the first section of this paper. One generally cannot make appropriate inferences when they are required, and one cannot readily tie the new information in the text to knowledge structures. This comprehension may require the construction of new schemata.

Chi (1976) has shown that background knowledge (rather than age) has a considerable effect on short-term memory capacity. Extensive background knowledge allows one to store whole chunks of data in a single encoding, rather than storing each element separately, by simply pointing to an already possessed schema. This may have implications for assessment of reading comprehension. For example, perhaps lack of background knowledge could affect performance on multiple-choice questions more than on open-ended ones since there is a problem of keeping the stem in short-term memory while reading the alternatives. Perhaps when answering questions with the text not available for referral the problem is more accute than when the text is available.

Such quantitative differences in background knowledge could, in fact, readily be assessed, particularly given the methodology presented by Anderson and Freebody (1979; see the section on methodology). Armed with such knowledge, we might be able to make certain inferences about the causes of reading comprehension failure for individuals. These deficits might also show up in inference questions. However, without the information specifically on prior knowledge, one would be unable to conclude whether the failure was due to a lack of prior knowledge or failure to use available prior knowledge.

Vocabulary knowledge differences and the reasons for their effects of producing discrepancies on standardized reading comprehension tests have received considerable attention (Anderson & Freebody, 1979; Hall & Tirre, 1979; Raphael, Freebody, Fritz, Myers, & Tirre, Note 8). These researchers have found extremely large differences in vocabulary between good and poor readers and across ethnic groups. If children have to expend much of their cognitive capacity identifying unfamiliar words in the test, less remains for building a model of meaning.

It is likely that lack of familiarity with structure creates problems just as does lack of familiarity with content. If the only type of material a reader is exposed to is narrative, it seems unreasonable to expect him to be facile with other types of text. This is so because familiarity with the structural cues allows the reader to find information more readily and know where inferences are required and what type they should be. A child who has never read a newspaper would have a good deal more trouble understanding newspaper items than a child who has had considerable exposure to them.

It is probable, however, that these differences are only the tip of the iceberg. The schematic knowledge and macrostructures themselves may be similar in extent but may differ considerably in their nature. For example, background knowledge for "police" may be similar in quantity for inner-city and suburban children, but differ qualitatively.

Qualitative Differences

The relationship between the text and prior knowledge can also be considered in terms of a qualitative mismatch between the text and the reader's background knowledge. This is a far more insidious problem. While quite subtle it could cause the reader to build a completely inappropriate model of text meaning without even being aware of the problem. It is not that inferences would not be made, but that inappropriate ones would be made. This problem could easily be self-compounding in that once the reader has begun to construct an inappropriate model, inappropriate inferences would be generated by virtue of the content of the growing model itself and the altered concept of "good form" so engendered.

Various contexts and instances of this type of mismatch are currently under scrutiny. For example, Hall (1977) has shown that with adults, social class and race, normally powerful predictors of reading comprehension performance, become less influential when content is made more job-relevant. However, even this does not completely eliminate the problem of racial differences. The developmental literature, too, indicates that children often can do a task if it is functional even though they may seem unable to when it is not functional (Yendovitskaya, 1971).

Perhaps we need to actively test for this type of mismatch, rather like testing inferencing versus eliminating it from reading comprehension tests. Washington (1979) found that minority students and nonminority students differed primarily in their performance on scriptally implicit test items. These are items which require the use of the reader's background knowledge in combination with the text. If our sociocultural knowledge would allow us to do so, we might construct specific background knowledge items which would indicate the presence or absence of the background knowledge required for answering the scriptally implicit items. We may find that such information provides us with a context within which to interpret performance on other parts of the test.

Cultural differences in the nature and understanding of nonliteral language similarly present a problem, since most of these depend on social conventions of communicative interaction and congruence of attribute salience in order to have their effect (Ortony, 1979).

The problem may well be even deeper than this. Hall and his colleagues are developing a mismatch hypothesis which suggests that not only does background knowledge differ across cultural groups, but their strategies for acquisition of knowledge also differ. The theory is that parents from different cultural groups reinforce different information-getting strategies. Thus, different predominant reasoning patterns would develop, and hence, a different answer on a comprehension test may seem correct. For example, "main-pointness" may be a more white middle-class phenomenon than "detail elaboration." Recent studies have hinted that inner-city black children may find temporal material relatively easy to comprehend, but have trouble with "compare and contrast" material (Hall, personal communication).

The student's cultural vocabulary differences may perhaps be controlled for by controlling the language in the test text. However, this will not alleviate the more general background knowledge problem. Several researchers (Royer & Cunningham, 1978; Tuinman, 1974; Washington, 1979) have suggested that making questions content-specific will solve many of the cultural problems. Unfortunately, this would probably require the use of only text explicit questions, and even these will often be affected by background knowledge (Johnston & Pearson, Note 7).

Asher (1978) found that the child's interest in the text (highly related to prior knowledge) can have a considerable effect on children's performance on reading comprehension tasks. Given the material found in most reading comprehension tests, however, it is clear that this has not been seriously considered in test development. Indeed, with group tests one can never completely deal with this problem except, perhaps, in interpretation of test performance.

The implications of the findings for reading comprehension assessment seem to be as follows. There are three possibilities for dealing with the effects of background knowledge in assessment:

1. careful selection of texts and questions so as to eliminate those which might contain "biases";
2. assessing in language which is appropriate to the reader's subculture as well as in the standard language;
3. assessing in various contents and language structures but including assessment which will discriminate between background-knowledge-induced problems and others.

Since the reader is likely to have to deal with a wide range of texts in the real world, assessment within a narrow range of texts for which there are minimal "biasing effects" of different prior knowledge seems of less interest than a more generalizable assessment. It does, however, seem unfair if such biased passages are included but not tagged as being such, as would happen in the case of the usual total scores which are used.

It is this writer's contention that we should be actively interested in examining background knowledge differences as sources of reading comprehension problems, rather than trying to avoid them. We are developing technology, in terms of assessment techniques (e.g. Anderson & Freebody, 1979), and the knowledge of specific areas which are indicative of such differences (Gearhart & Hall, 1979) which will enable us to produce potentially quite accurate measures of prior knowledge differences. Indeed, in informal assessements, teachers often ask this type of question already, although they have not been putting the information to full use in interpreting the remainder of the assessment because of a lack of understanding of the full import of the information.

The Task

It is extremely important that we should know exactly what we are asking the student to do when we set an assessment task. What factors other than that which we are interested in measuring will influence his performance? Research indicates that various reading comprehension assessment tasks can be considered in terms of the following factors:

1. production requirements;
2. memory and retrieval;
3. reasoning;
4. motivation;
5. purpose;
6. social setting and interaction;
7. expectation and perceived task demands;
8. test-wiseness.

Production Abilities

Production problems are difficulties in expression of information. For example, a child who is a very good reader may be poor at accessing information, particularly in an organized manner. Expressing ideas and organizing information from memory are skills which are not normally taught as part of reading, and skill in them may be quite independent of skill in reading comprehension.

Item types which are at the extreme on this dimension are free recall, which places a heavy demand on the students' oral or written production skills, and selection items (true/false or multiple choice) which make little demand on production skills. Other measures, such as, short-answer questions and underlining, lie in between these.

These differences could cause a problem in interpretation of performance on certain test types. For example, children for whom writing is a considerable chore, will doubtless have different criteria for what is worth writing down, from those for whom writing is simple. Production problems are almost certainly interrelated with retrieval, too, since organized expression may well act as a cue for recall of further information.

Memory and Retrieval

Selection items call upon memory in a different way from the open-ended items, and also require a different memory search strategy. Multiple-choice items can call on short-term memory for comparison of alternatives if the assessee uses certain strategies such as pairwise comparison of alternatives. Naturally there is a large difference in demands on memory if the source text is present or absent during question answering, though this difference may vary across item types.

Brown and Campione (1980, p. 9) summarize the retrieval problem as follows:

> In short, we have a great deal of evidence that: a) people frequently store information that they are unable to retrieve; b) the presentation of appropriate retrieval environments leads to access of material previously ´forgotten´; c) different testing

situations provide different retrieval environments, and therefore, assessments of the availability of knowledge varies as a function of retrieval support in the testing context; and d) the compatibility between encoding and retrieval contexts is vitally important as a determinant of the ability to access previously stored materials (Bransford, 1979; Norman & Bobrow, 1975; Tulving, 1978). All these arguments concern the optimal conditions for making information in memory accessible when needed; it is not sufficient to simply store information, for unless it can be activated when needed it is of little use.

Reasoning

There are several ways in which the task may require reasoning on the part of the examinee. Multiple-choice items may require more or less reasoning depending on (a) the strategy which the reader adopts to answer the alternatives and (b) the plausibility of the distractors. Open-ended questions can require reasoning out which possibility the tester wants from the range of possible self-generated alternatives.

Reasoning also enters into the picture via "test-wiseness" such that part of the task of the student is to figure out the alternative cues for answers (e.g., information in the stems of earlier questions, or two equivalent distractors which must consequently both be incorrect). Indeed, the strategies used by students in answering multiple-choice questions are important determinants of the cognitive demands of the task facing them. Reasoning ability may also be more important for less coherent text.

Motivation

Work by Asher (1978) indicates that the child's interest in the text content may motivate his performance, and sheer motivation has been shown by Lahey, McNees, and Brown (1973) to have a considerable affect on the test performance of certain children. The motivation may also be affected by peer group pressure (Labov, 1972) and the child's cultural background (Clay, 1976).

There are developmental differences in aspects of the text which readers find interesting. Stein and Glenn (1978) found that because first graders are more interested in the consequences of actions and fifth graders in the goals of characters, there are differences in what they consider to be the main point of certain stories.

Nicholls and others (e.g., Nicholls, 1979; Weiner, 1972) have shown many of the unmotivating effects of "learned helplessness" and it seems certain that this component of behavior will influence reading test performance, especially in the extent to which children will apply systematic strategies and will persist with difficult tasks. It has also been pointed out by Keogh and Margolis (1976) that children with learning problems tend to be field-dependent and impulsive. Impulsivity would be a particularly debilitating problem when dealing with multiple-choice tests. Nor should the problem of anxiety be overlooked. However, it appears that the reduction of anxiety in test situations must be accompanied by treatment of deficient reading and study habits which interfere with test performance, in order to be effective (Spielberger, Anton, & Bedell, 1976).

Purpose

The purpose for which people read a text tends to affect the types of information which they get from the text (McConkie, Rayner, & Wilson, 1973). Selection of reading strategies must be based on the reader's purposes for reading and these purposes can be controlled to some extent. They can even be manipulated implicitly via the types of questions which get asked in successive passages or within the text (Reynolds & Anderson, 1980). Currently, those children who know the strategies which are most productive in reading to answer multiple-choice questions have a considerable advantage in reading comprehension tests. However, those children who use such a strategy out of that specific context may find themselves at a disadvantage, since multiple-choice test items often tap details which would for many purposes be considered quite irrelevant.

The Social Setting and Interaction

Steffenson and Guthrie (1980) show that the language samples gained from questioning in a formal assessment/interview situation are qualitatively different from those gained from the same questions in a more meaningful situation. Harste and Burke (Note 9) found that when students read to them from a book and then gave them a free recall, the recall protocols were quite weak. However, when the children gave such a recall to a friend who had not read the passage, the recalls were quite full. Apparently, children are sensitive to the distinction between a socially productive and a contrived linguistic interchange.

Spiro (1975) found that with laboratory memory experiments subjects tended to differentiate stimulus materials to be remembered from other knowledge. That is, subjects concentrate on accurate verbatim recall instead of processing the input as they normally would, and they know that the information will probably be of no further use outside of the experiment. Hence, they severely reduce schematic integration. This is probably similarly true of assessment situations.

Expectation and Perceived Task Demands

Part of the reader's task is to figure out exactly what the examiner really wants to know or to hear and what he is up to. Consequently, readers probably respond to the same instructions in different ways. The effect of different variables is very much influenced by their perceived roles in the task structure. Consequently, instructions must be very clear and explicit, and even then the possibility of misinterpretation exists. Otto, Barrett and Koenke (1969) had children identify the main idea in four-sentence passages using the instructions, "Make up just one sentence in your own words that says what all the sentences tell you." Only 29% of their second graders could perform the task adequately. Danner (1976), however, found that instructions to find "the one thing that the sentences in the paragraph tell you about" enabled all of his second graders to get at least 66% of the main ideas. Danner also used orienting tasks to assist in providing task clarity. In these two examples, the experimenters intended a similar task but the children perceived different tasks, and rightly so. One task is asking for the selection of the topic while the other is asking for the construction of the topic.

Frederiksen (1975) used repeated read and recall with expectation of (a) recall or (b) application to a problem. Expectation of application affected the extent of inferencing and integration evident in the recall.

Tester control of the perceived task demands is small relative to the strength of expectation derived from precedent. However, one way around this is to first provide practice text and items to establish the task demands. Reynolds and Anderson (1980) have shown that when readers are repeatedly faced with questions of a specific type (technical term versus proper name), they perform better on those items, and concentrate more on and spend more time on the relevant segments of text. In any event, one should not leave the task ambiguous, since one would then have no idea what strategies different readers used for different purposes, and in what context to interpret the results of the test.

Test-Wiseness

Readers differ in their test-wiseness, that is, in the extent of their knowledge of and ability to use different test-taking strategies and alternate information sources which are available in the assessment context (Sarnacki, 1979). Millman, Bishop, and Ebel (1965) provide an outline of test-wiseness principles which assist test-takers in their efforts. Furthermore, it has been shown that training examinees in these skills improves their performance on subsequent tests (Slakter, Koehler, & Hampton, 1970; Wahlstrom & Boersma, 1968).

In the assessment methodology section of this paper, this type of task analysis is applied to certain assessment procedures in order to point out the information which they do and do not give to the teacher.

Summary

This section has presented issues relating to the importance of the structure, content, and language of the text for a reader's comprehension of it, particularly in the context of his background knowledge. The extent and manner in which these factors influence a student's performance in various contexts has also been described in terms of the percieved and actual demands of the task. An attempt has been made to point out some of the interrelationships between these factors and specific characteristics of the reader and the task which can affect assessment outcomes.

This section on the factors which affect reading comprehension and its assessment has provided input for both what to assess and how to assess it. These implications will be summarized and presented in the final section of the paper. The next section will address the issue of assessment methodology, and how some of the variables so far discussed influence it.

PURPOSES OF ASSESSMENT

Reading comprehension assessment is merely a more or less systematic sample of reading behavior which has been gathered for the purpose of informing a decision or statement. This decision or statement can be very specific or rather general, and the level of specificity carries with it certain implications for the constraints which operate on and the characteristics which we desire of the test.

There are three main levels of decision, each with its own limitations and freedoms.

Administrative

This involves decisions about large-scale resource allocation, or statements describing a given situation. It may involve looking at subgroups of the population for differential funding or evaluating large-scale programs. While this type of decision does not require such specific information as, say, diagnostic assessment, it has often been accused of failing to pick up important differences because of the global nature of the assessment.

The latitude which this assessment allows is that all students do not need to give the same sample of behavior. That is to say that one can use a very large test which has been divided up amongst students, only some of them having the same items. This is because decisions are at the group level rather than at the level of the individual student. This level of assessment must be static, in that there can be no individual response-dependent adjustment of assessment methodology or difficulty. The test will be administered to students of a wide ability range and must have items of appropriate difficulty for the whole range.

Diagnostic

This type of assessment involves making instructional decisions at the individual level, selecting materials, and teaching strategies. The requirement here is specificity of information. However, for several reasons there is more room for flexibility here.

Firstly, since the assessment is not normative but ipsitive in nature, assessment can be individualized. That is, one student may have a completely different set of assessment materials than another student, especially since one does not require the spread of difficulty which is necessary for group assessment. Secondly, the assessment can be dynamic, in that the technique, the level of difficulty, and the content can be response- and person-dependent. There is no requirement of standardization because if the domain is specified clearly enough, and the specific skills are assessed adequately, the test can be criterion-referenced. Neither is there a requirement of always examining a specified number and pattern of texts and tasks. Provided that the assessor knows what is being required of the child and what it means in the context of a theory of cognitive functioning, the tasks, texts, and contexts can be varied so as to address the hypotheses which the assessor is shaping about the child's difficulties.

Selection and Classification

These are generally group-administered tests which are uniform so that all students are compared on the same basis. These currently have the greatest limitations in that the test has a time limit (as do the others), but it must still cater to a wide range of pupils, and each pupil must generally get the whole test. That is, we have a test which is not diagnostic or sensitive for individuals, yet we wish to make individual decisions on the basis of it. These limitations are rather severe in their effects on the standard pencil-and-paper tests.

While this is generally the case, it need not necessarily be so. For example, provided the domain is well enough specified, computers could be used to skip segments of text and tasks which, given performance on earlier sections, would clearly be too easy or too difficult for a specific student.

ASSESSMENT METHODOLOGY

Having achieved some consistent idea of what we are attempting to assess and why, and a good idea of which variables are likely to have an effect on an individual's performance, we are ready to consider how we might assess reading comprehension.

It would be very nice to have a procedure which is practical (in terms of fitting it to the constraints under which it is likely to be used), measures what we want it to measure, and measures only what we want it to measure. However, apart from the obvious problem that most reading comprehension processes are not directly observable, there are many constraints which affect the nature of reading comprehension assessment.

Constraints upon Assessment

Practical Concerns

A major constraint which invariably operates is time, and this is so for a number of reasons. Those working in the field, especially those in the classroom, have minimal amounts of time to devote to testing, especially individual testing, and in any case, the examinee cannot give full attention for prolonged periods of time.

The longer the delay between reading and testing a piece of information, generally, the more difficult it is for the reader to access it; thus, sometimes requiring a different score interpretation. The more test items there are, the more chance there is of inter-item contamination. On the other hand, the longer a test, the more reliable it tends to be statistically. This presents a real problem for assessment, especially that which attempts to make more broad generalizations about an individual's reading comprehension ability. Normally, a compromise is struck so that a brief group self-administering test is given, and an attempt made to gain maximal information from it by including in it a large number of short passages each with a few questions.

The requirement of machine scoring of mass testing for administrative purposes also places severe limitations on test developers. Currently, at least, items are restricted to the selection type (true-false, multiple-choice, matching).

Item Sampling and Generalizability

Much sacrifice has been made in order to meet the above demands, but even more has been made to meet the demands of sampling theory. The idea in classical measurement theory is that one can generate an infinite number of items about a passage and since one cannot administer all of them to a single examinee, one should administer a random selection of them. One of the upshots of current research is to demonstrate that, with reading at least, this is a very dubious approach (Anderson, Wardrop, Hively, Muller, Anderson, Hastings, & Frederiksen, 1978).

It is now very difficult to argue that a random sample of questions will give us effective information, since it is clear that certain information in the text is more important than other information, and that sorting one from the other is itself an important comprehension ability which cannot be taken for granted. As many have pointed out before, a systematic method of item development (preferably rule-governed) would be very useful. Even a rule-governed procedure for selecting items of specified types would be a great help. It seems that the prerequisites for this are beginning to accrue.

Such an approach might profit from the research on text structure as discussed above. If we can develop formalized ways of representing the information which is "in" the text, the interrelations between information segments, and the structural levels of importance, this may be used to select items in a more systematic manner. We could sample the kinds of information which we really would like to know about. If we can classify the type of information or relationship which an item is tapping, then we can begin to consider how the reader is handling various intratext relationships and perhaps generalize within these across texts or reading goals. This knowledge will also help us to define the domain to which one can generalize the outcome of a particular assessment device (Anderson et al., 1978).

The van Dijk (1977) and Brown and Day (Note 4) work on rule-governed summarizing presents a systematic method of stage-by-stage reduction of information to the most important. These rules could potentially be used to generate selections of questions at each level of importance or to select questions on those grounds. However, here it must be conceded that, as discussed above, importance is also governed by the perceived task demands and goals of the reader. Such systems would thus presuppose congruency of reader and author goals.

These systems also presuppose some sort of rational procedure for specific item generation and interpretation. Furthermore, as pointed out above, the passage from which the questions were derived can no longer be considered a random factor either, thus making generalization across texts difficult. This is especially so when the reader—text interaction is considered, at least in the individual case, where one cannot count on the text tapping a "representative sample" of the individual's background knowledge.

What may be required are more specific tests. One possible decision point which arises from the work of Hall (1977) is the functional aspect of the passage. That is, one should be concerned about choosing passages which have potential functional value for the reader -- thereby ensuring attention and ecological validity, and minimizing bias. However, in doing this, one must be careful to assess possible increases in the passage independence of the items, since the more functional or interesting a text is for a reader, the more likely it is that the reader has some relevant background knowledge.

Passage Dependence

The whole problem with the item-sampling approach is that one does not really wish to know about the recall of random pieces of the text. The real interest lies in the degree of integration, inferencing, and general tying together of the textual information with information possessed.

With regard to assessing this latter aspect, it would be interesting to have some measure of what background knowledge the reader had before reading the text. Without this it is difficult to estimate his integration of new information with the old.

This argument leads us to reexamine the old issue of passage dependence of comprehension test items (Farr & Tuinman, 1972; Hanna & Oaster, 1978; Pyrczak, 1975-76; Tuinman, 1974). Test items are considered to be passage-independent if reading the passage makes no difference to performance on the item. Items which have been assessed as being passage-independent have been frowned upon for some time, yet without such items, we can have no idea of the extent to which the individual has integrated the new knowledge with his background knowledge, nor any idea of how much background knowledge he had in the first place.

It is here that the major cultural bias issues become apparent. Biased items are normally statistically identified and eliminated from tests without any idea of why they are biased. These items are most likely to be items which involve some inferencing (Washington, 1979). This is where background knowledge enters the picture. Cultural differences in background knowledge can cause different inferences to be made. Thus, any item whose answer is not explicitly stated in the text (and even some of these, since conceptual knowledge of single words can differ) is subject to possible bias due to the differences in background knowledge (the source of the inference). Thus, these items tend to be the passage-independent ones. To eliminate all such items would leave largely what have been called "low-level" questions.

One possible way of dealing with this problem is to include some of these passage-independent questions deliberately. By doing this, we could perhaps explicitly tap the requisite background knowledge. This could allow us to begin to answer such important questions as why certain items are biased, and why certain individuals produced the pattern of responses they did. We can ask whether the reader had the requisite background knowledge and whether or not he used it, or how it differed from one subgroup to another.

Research has shown that passage-independent items (a) measure something different from passage-dependent items (Tuinman, 1974), and (b) if properly designed, provide a good measure of background knowledge (Johnston & Pearson, Note 7). At issue, then, is a new approach to passage dependence, shifting from a statistical definition to a conceptual one.

The practicality of this possibility is given added weight by the observations of Anderson and Freebody (1979). These researchers, in noting that tests of vocabulary knowledge correlate very highly with tests of reading comprehension, suggest that a plausible explanation is that the word knowledge merely indicates a stronger prior knowledge of the area. If this is so, then a few well chosen (with respect to specific sociocultural differences), passage-independent vocabulary items could possibly tell us quite a bit about the reasons for certain aspects of an individual's (or group's) reading comprehension performance by indicating that the test was approached from a culturally different perspective.

Item Generation, Selection, and Classification

In the past there have been several attempts to formulate algorithms for the generation of reading comprehension questions (e.g., Bormuth, 1970). These algorithms were based on largely syntactic manipulations of the text. The rationale behind attempts to derive question-generation algorithms is that in current tests of reading comprehension, the questions used are based upon the test developer's whims, and different developers will certainly produce different questions from the same text segment. Consequently, there is likely to be limited agreement in the specific conclusions they reach. Elimination of this capricious element is desirable (but not to the extent that it leads to stilted or ecologically invalid assessment, as many other attempts at increasing reliability have done).

It would be a very good situation if we had certain systematic and theoretically reasonable rules or algorithms by which one could generate appropriate questions for a given text-reader-task situation. Unfortunately, even attempts to generate such rules for specific texts have so far been unsuccessful. There has been, however, more success for post-hoc analyses which take the developed questions and analyze them with respect to certain of their characteristics.

Much information can be gained from a theoretical analysis of the relationship between the question asked and the source text. Lucas and McConkie (1980) have looked at questions in these terms and developed a system for classifying the questions. The system involves a propositional analysis of the text and a subsequent analysis of the source of information required to answer the question.

Pearson and colleagues (Pearson & Johnson, 1978; Raphael, Winograd, & Pearson, in press) have attacked the same problem from a different perspective. Their approach is to examine the relationship between the answer to a question and the text.

The two approaches overlap considerably. For example, one type of question relates to the ability to combine information which is explicitly stated in different parts of the text. Pearson and Johnson (1978) call these textually implicit questions. In general, it seems more difficult to combine information from two propositions which are further apart than from two that are close together. Similarly, it may be more difficult to combine the information in five propositions than that in two. Lucas and McConkie (1980) have included this information in their classification system. They refer to this type of question as "Inferred" and code the numbers of the relevant propositions in the text (which has previously been propositionally analyzed).

Examining the processes involved in these questions, and even coding them, is complicated when information is repeated or is implicit in several places in the text. This type of question appears to tap "reasoning in reading"--essentially logical reasoning skills such as solving syllogisms (if A and B, then C)--and shows the close proximity of the work on problem solving presented in the first section of this report. It may well be that this type of skill is, as in problem-solving, readily diagnosable and readily instructionally rectified.

On the other hand, these item type descriptions may only be applicable when the reader has text and questions available at the same time. Once the information is stored in the head, little of the surface structure of the text remains. Thus it may be that these item-types lose some of their meaning once the text is no longer accessible. What is more likely to remain is a trace of the macrostructure of the text in the structure of the stored knowledge.

Perhaps these item classification approaches can be further developed in the light of the findings cited elsewhere in this paper. For example, as yet, the question classification systems do not take into account such things as the staging of the information. Beck, McKeown, McCaslin, and Burkes (Note 10) have proposed that questions about narrative text should be based upon the Narrative Analysis (Omanson, 1979). This approach would allow a fairly objective classification of questions relative to their "centrality" to the main strand of the narrative. The centrality of information to the causal chain in narrative has been shown to have a strong effect on recall (Omanson, 1979), as has the staging of the information in expository material (Meyer, 1975) and whether the material is hierarchically or temporally organized (Thorndike, 1977). These factors may turn out to be more important determinants of performance than the components currently involved in the taxonomies. Research is needed in this area.

As a further example, the "main idea" type of question at first appears to be an extension of the "implied" type of question. However, the work of van Dijk (1977) and Brown and Day (Note 4) has shown that certain other explicit knowledge and skills can be utilized. Brown and Day have shown that children who are not good at summarizing text can be taught the skill by supplying them with knowledge of explicit summarizing strategies. Perhaps in the text-available situation the "inferred" question is only assessing a subset of the strategies involved in main-point finding.

As a further example, the long recognized "inference" or "scriptally implicit" (Pearson & Johnson, 1978) question, if looked at in terms of Warren, Nicholas, and Trabasso's (1979) classification of inferences, may decompose into a series of more specific item types which are also more relevant to the manner in which the knowledge is stored some time after reading. Given Omanson, Warren, and Trabasso's (1978) finding that inference probes are the most sensitive indicators of reading comprehension, a carefully designed set of such items may give us quite detailed information about processing at various levels of reading comprehension, from the single word level to the discourse level.

Unfortunately, the work that has been done so far only deals with the stems of questions. There has been only limited application to the alternatives of the common multiple-choice item. This is a crucial next step if we are to continue to use these items, since the nature of the alternatives in part determines the cognitive demands of the item and, hence, the information which it yields.

Since the psychological reality of the systems of classification has not yet been established, much research is needed in this area. Clearly the approaches have important implications for the development of tests. For example, multiple-choice items could be based on Lucas and McConkie's or Pearson and Johnson's type of analysis, so that stems would have a systematic relationship to the text and to each other. The inclusion of staging or causal chain information would at the very least provide operational definitions of main-point and detail questions. Response alternatives could be based, for example, upon Spiro's (1980) analysis, so that production of different response options could be based on different, identifiable cognitive demands. Should such a procedure work, then diagnostic information could be gained from the test (subject to the reservations presented in other sections of this report). The diagnostic value of the test would stem from patterns of responses on similar item types which would point to background knowledge or text cues which were or were not being used (appropriately or otherwise). Patterns of responses across similar types of alternatives might lead to a diagnosis of the kinds of cognitive difficulties which arose. The kinds of difficulties which might be picked up are similar to those outlined by Spiro (1980; see the first section of this report).

What one concludes from the foregoing discussion is that there are certain dimensions along which questions might be classified. The first is the source of the information which is required--text or background knowledge (or both) and possibly location in the text, especially in terms of its macrostructure. These dimensions may differ in their importance with the availability of the text for reference when responding to questions. The second dimension concerns the cognitive demands which the question places upon the learner. Classification systems should also specify the relationship between the classification system and various reader purposes. In order to develop this dimension, there is still a need for research on the relationship between specific reader purposes and relevant reading strategies.

While it seems that we are still unable to produce effective algorithms, we are approaching the stage of being able to classify items and item clusters with respect to the information which they could yield. Thus we approach a position from which to select items which have a clear relationship to the structure of the text, the reader's prior knowledge, and the nature of the requisite cognitive processes. Knowing the characteristics of these item clusters, we should be able to generate tests which provide more, and more meaningful, information.

Reliability and Validity

There are two especially important aspects of validity in tests of reading comprehension. The first is concerned with the meaningfulness of the construct with which we are dealing. This has been a big problem for diagnostic tests, since the reality of the individual "subskills" being tested has been suspect. Our increasing understanding of the processes involved in reading comprehension will enable us to more rationally evaluate this aspect of the test validity.

The second aspect of validity refers to the degree to which our chosen assessment technique actually reflects ability in the specific skill which we claim it measures. To this end, some examples of the demand characteristics of various tests and test situations appear at the end of this section of the paper. Careful study of these areas is the road to a rational understanding of this aspect of validity.

Closer understanding of the relationship between reader and text, and of the nature of text itself can be used to increase the validity of standardized tests, since we are beginning to know more about what processes we are measuring with certain texts and item types. Knowing the strengths and limitations of the different techniques may yet enable us to combine information from different measures to give us a highly valid measure of exactly what we wish to assess. Furthermore, since invalidity lies as much in the claims which we make about what we have measured as in the instrument itself, understanding the strengths and limitations of the instruments should help us to make more valid use of the tests.

Reliability is also a very important aspect of measurement instruments. This refers to the consistency of the measure across different measurers and occassions. It has generally been argued that standardized tests have increased reliability but have achieved it at the expense of a certain amount of validity, especially ecological validity (the naturalness of the task). Informal tests, it is claimed, have the reverse failing, reliability being quite low (especially between scorers) but validity being higher. To an extent this is true, but progress is being made toward reducing the problem. For example, the development of text grammars has enabled more reliable and meaningful scoring of free-recall protocols--scoring which also seems to have some diagnostic validity. Here again, however, one is faced with the problem of deciding on the source of difficulty, i.e., input, retrieval, or expression.

Reliability should be looked at in terms of interpretation as well as scoring of tests. This is the intersection of reliability and validity. Reliability is a prerequisite for validity. The context in which a test is administered and certain demand characteristics of the text which are not integral to what we are attempting to measure are extraneous sources of variability in test performance; they are a threat to reliability. These factors may also invalidate an instrument which under certain circumstances may be quite valid. For example, the demands of the social situation in which assessment occurs can strongly influence the outcome of the assessment. This is clearly shown in the study by Steffenson and Guthrie (1980) in which the same task (the Peabody Picture Vocabulary Test) in different social contexts (test situation versus a real sorting problem) produced different results. The reasons for such differing results have been pointed out in the earlier section on task analysis. Differing results, of course, would make the measure unreliable across these situations. Knowledge of these situational and demand characteristics will help us deal with this problem. For the present, however, there are problems with our current concepts of reliability and validity as they relate to tests of reading comprehension.

The concept of reliability, as Tuinman (1979) has pointed out, forces test makers to ask a large number of questions, many of which must test trivial information. This is so for four reasons:

1. The desire to generalize across different samples of text structure, content, and difficulty, has led test makers to include a number of text segments, which, because of time constraints must necessarily be short.
2. Questions should be independent, and with a short passage, one is put in a position in which detail questions are very tempting.
3. Sometimes there is an attempt to use a number of questions of the same type in order to get reliability for each item type.
4. Items are selected on the basis of high discrimination indices, and the items which are best at discriminating are often those testing trivia (Tuinman, 1979).

One of the upshots of this set of constraints is to ensure that the brief passages are stilted, since the writer must cram information into each one so that a reasonable number of independent questions can be asked. Thus, the type of information which is often tested is knowedge of detail, which, for many reading purposes, is quite unimportant. We are consequently stuck with a measure of what children learn when they read for a specific, often dysfunctional, purpose, i.e., to answer multiple-choice questions. Worse still, children seem to have come to expect this of tests, so how can we effectively assess their performance when they are reading for other purposes?

It seems that in our efforts to gain reliability, we have sacrificed validity. This has resulted from a lack of understanding of how to evaluate the test other than statistically, and an attempt to answer in a single test the rather vast question, "How capable is this child, with respect to his peers (including all subcultural groups), at reading comprehension in all types of text, and across all content and all situations?" Given the constraints on assessment, the question is one in which only administrators and parents would be interested. Since the parents in particular would be generally unaware of the limitations of the assessment, their interpretation of such numerical information would be more or less invalid, especially since their interpretation would generally be based on a single score. Rather than such mean performance information, the classroom teacher is more interested in individual differences which can aid in instructional decision making.

A line of research which is thus of particular interest concerns the "zone of proximal development" (Vygotsky, 1978; Brown & Ferrara, Note 11). This research is concerned with the inadequacy of current static assessments of children´s I.Q. The argument is that children with equally low I.Q.´s are not necessarily of the same learning potential, the difference becoming apparent when an adult or more able peer provides some assistance. It is the extent of improvement which occurs under some optimizing procedure which constitutes the zone of proximal development.

Such research has attempted to shift testing away from a standardized approach which attempts (but often fails) to hold context constant across examinees. The preferred direction is toward using the interaction between tester and examinee to find the conditions under which certain tasks can be performed, and how much of specified types of assistance is required for the child to succeed. In terms of reading comprehension assessment, this may amount to setting up different text and task situations which are apparently just beyond the reader's skill, and providing increasing amounts of information and instructional assistance. This has the potential of being considerably more diagnostic than other approaches, but its optimal use rests on the development of a set of texts and tasks with an associated sequence of prompts which have diagnostic value. For example, one could gradually reduce the demands upon production skills, organization, memory, and memory access by providing different tasks, and increasingly blatant prompts. Some of these prompts could be strategic in nature and others informational.

Thus, it seems neccessary for us to shift our perspective on reliability and validity, emphasizing the latter in the light of the factors which influence reading comprehension. Increased validity should incidentally improve the reliability of our interpretations in particular. This emphasis on validity will also force us into a different perspective on generalization, since it is clear that we cannot simply be random about text, tasks, or context selection and ensure generality. It may be that we should be more interested in specifics than in generalities. Trying to reduce the effects of different factors may not be the way to do it either. For example, attempting to remove the effects of background knowledge by eliminating passage-independent questions simply will not work.

The Cognitive Demands/Information Content of Assessment Methods

As pointed out above, since reading comprehension is a mental activity, it is only available for indirect, second-hand scrutiny. We can never actually watch the mental operations, but must infer them from other sources of data. In making these inferences, we must be very clear about the grounds we have for doing so. In order to be so informed, we should understand (as clearly as our data and theory will allow) the actual demands and assumptions involved in our assessment techniques. For example, the process and product techniques are supported by different sets of assumptions.

Most assessment of reading comprehension has to date been of the product type and consists of a series of short passages (sometimes less than ten words) each followed by multiple-choice questions which bear some, often ill-defined, relationship to the preceding passage. These measures represent a very restricted view of reading comprehension, especially in their current state of development.

The object of much past development of reading comprehension assessment has been a single test which will tell us all we want to know. It is our contention that pursuing this objective in the current manner is futile, given the complexity of the reading task and the number of variables to be assessed and/or taken into account. We believe that the more reasonable approach is to refine and use the variety of approaches to measurement which we have available already, in the light of our knowledge of the skills and abilities involved in each, though we might want to add some supplementary approaches. By appropriately selecting combinations of these measures, we may gain a more clear picture of what and how a reader comprehends under a given set of cicumstances.

Product Measures

Free recall. The most straightforward assessment (in terms of initial preparation) of the result of the text-reader interaction is a free recall. However, it becomes clear at the outset that ease of preparation of the measure is inversely proportional to the ease of interpretation. While many researchers (and their assistants) will attest to the amount of time involved in scoring free recalls, such recalls may tell us something about the organization of the stored information. In combination with other measures, we can also make some inferences about the retrieval strategies which the reader uses. This, in turn, gives information about probable long-term recall of the newly-gained information.

Recent research has enabled scoring of recall protocols in fairly meaningful and consistent ways (Frederiksen, 1975; Mandler & Johnson, 1977; Stein & Nezworski, 1978; Fredericksen, Note 12; Turner & Greene, Note 13). For example, story grammars have allowed scoring of stories in terms of the presence or absence of the integral elements of the narrative structure. Summarization strategies such as those described by van Dijk (1977) and Brown and Day (Note 4) may ultimately allow us to examine objectively the extent to which the individual has been able to assimilate and reconstruct certain types of text. Patterns of intrusions, distortions, and omissions may provide valuable information on specific influences of the individual's background knowledge.

Although there is a lot that one can infer from and about the things that _are_ recalled, one can say nothing about the comprehension or memory of what is not recalled. There are a number of possible reasons for failure to produce information in free recall. For example, production deficits are a problem. Alternatively, schema selection or access may be blocked or inappropriate. Or, as Harste and Burke (Note 9) point out, it may be that the reader simply misinterprets the task demands and, assuming (or knowing) that the tester has read the passage, gives only a cursory protocol.

What then are the cognitive demands of the free recall task? We must be fully aware of the large memory component which is invariably involved and the problems of retrieval. The reader must first clearly understand what is required of him in terms of the level of detail he is being asked to retain and reproduce (gist recall to complete recall including all pragmatic and possible inferences) and the degree to which recall should maintain the surface structure of the original passage. He must have understood and stored the information, and be able to retrieve it on demand. He must then decide on a point at which to start, and a path through the information, e.g., start at the beginning and procede to the end, recalling as much detail as possible, or give gist first, then return to the beginning and elaborate, etc. A decision must also be made on a perspective from which to present the recall (Bower, 1978), and this may involve hypothesizing about the tester's perspective and working from there.

Production skills (oral or written) are also required, and production expertise differs across individuals. Unfortunately, reading comprehension skills and production skills are not perfectly correlated. Consequently, failure on a production reading comprehension assessment task cannot be clearly attributed to production or comprehension skills separately. On top of this, the reader must be motivated to "play the game". Failure to meet one or more of these demands may cause similar recall protocols (in terms of the quantity and the pattern of the recall), so we must be wary of drawing conclusions such as: not-so-fluent readers remember less of the information, or the not-so-fluent reader does not remember information from discourse in an organized manner (Marshall & Glock, 1978-79) without gaining more information than that contained in free-recall protocols.

Probe questions. Probe questions following free recall locate more information which the reader has stored. Probes which present different perspectives are likely to produce different extra information (Anderson & Pichert, 1977), which is something that should be taken into account in free- and probed-recall assessment. Incidentally, such use of perspective as a retrieval strategy may well be worth teaching to students.

A problem with probes is that, to date at least, there is no systematic method of generating appropriate ones and ones which would be consistent from tester to tester or passage to passage. Another problem is that even using the probes does not exhaust what was learned from the text. A recognition test or sentence verification task (Royer & Cable, 1975) may tap even more information. The extent to which this is so is useful information in itself since it may tend to indicate a retrieval problem rather than a storage or encoding problem.

One line of research currently being pursued (McConkie, personal communication) is a method of systematic probe generation. This involves analyzing the text and then locating the highest-level node not recalled in the free recall. This node is then used to generate the probe and, after further recall, the next highest node is located, and so on. It seems necessary to computerize this procedure, since one would require a minimum of delay between reading and free and probed recall.

The use of probes raises a further issue. Is there a qualitative difference in the comprehension between the information that was freely recalled and that which required probing, or is it perhaps merely a retrieval problem? For example, what of the reader who can answer any question on what he has read but if asked for a free recall has a great deal of difficulty producing an organized response? It certainly seems that retrieval and production problems are common enough and are amenable to instruction (Bransford, 1979; Brown & Day, Note 4).

The big drawbacks associated with the free recall plus probes are the following: (a) For the teacher, it is very time-consuming to administer and score, and requires training and practice. (b) For the administrator, scoring of large numbers of recalls poses horrendous problems. These problems would exist even if the protocols were written, which is ill-advised because the less capable students tend to write less well also, putting them at a great disadvantage. (c) Generally, for younger and less able children especially, oral free recall is required. This must then be transcribed. If the most efficient probes are to be used, then they must be individually and rapidly computed. The probes can, of course, be used without the preceding free recall. They then constitute open-ended questions or the stems for multiple-choice, true-false, sentence-verification, or other items. (d) There remains the problem of separating the effects of input, storage, and retrieval difficulties.

Open-ended questions supply different information from the free-recall information. While they are likely to tap more information, they do encourage further processing of the stored information. For example, inferences which a good reader might make while reading may not be made by a poor reader until the probe suggests the value of making such an inference. This complicates the interpretation of any of the probe-type questions. One also runs the risk that information presented in one probe may affect performance on another. For open-ended questions the production problem is still present, but to a lesser degree than with free-recall. However, somewhat different operations may be required by different probes, depending on the extent to which the original surface structure is represented in the probe. This is, in fact, a valuable piece of information, since individuals may perform differently on the probe depending on the extent to which they stored (and were dependent on) the surface form of the text.

True-false questions. The next most straightforward type of question is the true-false question. These questions eliminate the production problem from the task demands but they suffer from certain other drawbacks. To begin with, a chance score is 50 per cent, which makes score interpretation rather awkward since the chance component is difficult to extract. Hence, one can never know why the reader gave the correct or incorrect response.

For the reader, there is a matching procedure required. If the surface structure is the same as that occurring in the passage, it may be a simple match. If not, then the reader must transform either his stored information or the information in the stem and attempt to match. If surface structure is similar to the original but the statement is false, then selection of meaning as being the more important cue is required and a belief in one's own ability to understand. There may be an approach-approach conflict. The meaning of the stem seems wrong, but the surface structure may seem right. It is likely that a poor reader, who has had considerable experience of his own failure to comprehend, may well understand but decide to go with the surface structure indication since that is more compelling evidence to him than his dubious transformation.

The low signal-to-noise ratio seems to make these items less useful as diagnostic tools. However, Anderson and Freebody (1979) used them effectively in estimating vocabulary size. Their strength lies in the breadth of material that can be sampled in a reasonable timespan and their validity as a gross measure is supported by those writers, though for a different purpose than reading comprehension. In order to correct for guessing, they embedded nonwords, which gave them an estimate of the student's false-alarm rate. This allowed them to use a signal detection analysis to correct for guessing. These may well prove to be functional for administrative or broad, descriptive types of assessment where one might want to make statements about the behavior of a broad sample of children over a broad sample of texts. However, one is still left with the problem of generating the stems for these items.

Multiple-choice questions. Most of the common reading comprehension tests use multiple-choice items. This item type is probably the most researched, most maligned, most difficult to construct, most abused, yet most functional of all items (when properly harnessed). It has the potential for reducing the problem which is evident in the true-false item, that of knowing why the respondent gave the answer.

The potential of the multiple-choice item for use and abuse is immense. The positive aspects have certainly not yet been capitalized upon because of a lack of theoretical and empirical research. Currently, the multiple-choice item serves only to allow mass objective scoring and to reduce the chance score from that of the true-false question. That is, we still get the same information (right versus wrong) as we got from the true-false item, there is just less noise. In fact, the multiple-choice item can give us much more than that. However, the extra payoff is dependent upon the development of a theory-based procedure for producing the alternatives. If selection of a specific alternative by a reader indicates a particular reading strategy or problem, then we can begin to look at patterns of similar responses, and the assessment can become more efficient. This legitimizes the questions "Did he get it right for the right reason?" and "Why did he get it wrong?," because one could examine patterns of responses to correct and incorrect alternatives and make inferences about the strategies being used.

While the multiple-choice item still suffers from the problem which afflicts all forms of probed recall, that the cuing can induce processing which would not otherwise have occurred, this item type contains certain safeguards against the problem: by supplying a variety of alternatives which suggest different processing strategies, respondents can no longer find unequivocal incentives towards specific extra processing.

The multiple-choice item in its common form probably involves a great variety of extra processing skills. There are specific strategies to be used. These need to be learned, and while they could be considered to present a type of reading comprehension, it is clearly not a type which we wish to be assessing in depth. These items require understanding of the stem, sometimes holding the stem in short-term memory, and using one of several possible strategies to evaluate the alternatives. Of course, different individuals will use different strategies, some of which will be maladaptive. For example, some readers do not read all the alternatives, they stop when they think they have the correct one.

Clearly these skills are of a problem-solving nature. Some are not especially necessary for reading comprehension, and we would rather that these did not enter into the assessment. On the other hand, we may be able to detect failures which are associated with a lack of these skills by using the information in the response patterns. Another way around this may be to allow the reader to estimate the probability of correctness of each individual alternative (Pugh & Brunza, 1975; Johnston & Pearson, Note 7). This has three advantages. Firstly, it guides the respondent through the steps of rational responding to the multiple-choice format. Secondly, it does not force the individual to guess, although if a final answer must be given also, the tester would have information about the guessing behavior built into the rating scales. This would provide continuous rather than dichotomous information about each alternative, so the possibility of diagnostic patterns emerging (assuming theory-based stem and alternatives are used) is considerable. A third possibility is that we go to some length to teach all children test-taking skills. This would still not completely eliminate the problem, since, as usual, some learn better than others. Yet another alternative is to allow space for the child to give an explanation of why an item might be considered ambiguous or have other problems.

Availability of text. Reading comprehension tests may be administered in two different ways, with a passage available for reference as the questions are answered, or with it unavailable. While these are often used interchangeably, the difference in the cognitive demands of the two approaches is considerable. Again, if we recognize these specific differences, they may become less of a problem. Indeed, we may use the contrasting information sources to provide us with greater depth of insight into the processing differences between individuals. Assessment with the text absent clearly produces greater demands on the long term memory, and cuing doubtless has a stronger function. Retrieval and organization skills are more important also. With the text present, more weight is placed on recall of approximate location of the information in the text, knowledge that one should look for it, search strategies, and logical reasoning skills. On an individual basis, the time taken and strategies used to locate information in text are also useful sources of information which can readily be observed. Summarizing text in text-present versus text-absent conditions represents the difference between the ability to summarize and the ability to store, retrieve, and organize and summarize information.

Process Measures

The process measures are more or less on-line assessments of what happens when one reads. Some of these contain certain amounts of product assessment in that they are dependent to some degree on memory components. Furthermore, while these measures are "on-line," they are still indirect indices of comprehension processes but in a different way from the product measures. They do not necessarily explicitly indicate that the reader has comprehended, only that certain behaviors are occurring during reading. Their use rests upon a different set of assumptions from those underlying the product measures.

Miscues. One class of process measure involves oral reading. While this may or may not be representative of normal silent reading (there is currently no consensus about this), it can certainly tell us something about how readers comprehend text under certain circumstances. One of these on-line measures which some teachers already use is the Miscue Analysis, initiated by Goodman (1968). The rationale behind this lies in a conception of the reader as a rational cue user. In order for the reader to comprehend text, his task is to make use of three cue systems: syntactic, semantic, and graphophonic. By comparing a reader's oral production with the text, Goodman argues that we can make inferences about the progress of the reader's comprehension. For example, if the story is about a stallion and one child reads station and another reads horse, one can infer certain differences in the state of their comprehension.

If reading comprehension is considered to be the systematic use of the cue systems to understand the intended meaning of an author, then Goodman's miscue analysis is a good basis for measuring certain dimensions of reading comprehension. It is, however, only a basis since, as is indicated elsewhere in this paper, there are many more potential cue systems that the reader might use, such as cues indicating which type of text one is dealing with (expository, narrative, legal document), cues indicating the task demands of the reading experience (social, personal, explicit, implicit), cues indicating the content area and depth (and hence the requisite extent of background knowledge).

The cloze task. The cloze task has been used to assess comprehension in many experimental studies and in classrooms, and as a benchmark for determining the readability of a text (the other side of the reading comprehension coin). While it is often used as a product measure, it is perhaps most informative when used as a process measure in oral reading. The task has its advantages. It is quick and very easy to construct in its basic form. One merely removes every _n_th word and studies what the reader does at the blanks. Unfortunately, it does not tell us why the reader does what he does, and the measure is not without its problems. To begin with, performance is related to literary style. Hence, even if normally one would have relatively little difficulty with a passage such as part of King Lear, in the cloze situation considerable difficulty may be encountered. This is especially so if the common "stringent criterion" is used. That is, if a response is only considered correct when it is an exact replacement of the original word. Such a system is probably often deceptive, at least for individual items, since one who understands well is quite likely to have available a number of synonyms. However, this may be less of a problem across many items.

The cloze task cannot be considered a normal reading task because often one must hold an empty slot in memory until one can locate information to fill it and construct a meaning for the segment. This places quite a demand on short-term memory, and there are search skills involved. On the other hand, if the characteristics of the text and the approximate level of prior knowledge are known, one might be able to define for a given reader situations under which specific problems arise. Thus, remedial strategies could then be taught.

A useful modification to the cloze procedure has been made by Yanofsky and her colleagues (personal communication). The modification involves deletion of specific categories of words in specific contexts rather than every _n_th word. Through careful selection of the type and environment of the words, one has more control over the nature of the task which the reader must perform in each case. This systematic knowledge of the syntactic, semantic, and graphophonic context at levels from word to discourse allows the assessor to put together information and make informed judgements of certain reader strengths and weaknesses. There is still much to be learned about the cognitive demands of cloze tests, and research effort should be directed towards this end.

Eye movements. There is a growing body of literature on the use of eye movements to infer the strategies which individuals use in the act of reading. Currently, the necessary equipment is expensive and confined to laboratories. Nonetheless, this might not always be the case, and there is evidence that this type of study will help us to pinpoint individual differences in reading strategies (McConkie, Hogaboam, Wolverton, Zola, & Lucas, 1979). A record is made of what the eye does (as it scans the text) in terms of the location and duration of fixations, and time, speed, and acceleration of movements (saccades). The frequency and characteristics of regressions (look-backs) are also recorded. Armed with this information plus a detailed analysis of the characteristics of the text, one can infer some of the underlying cognitive processes by considering, for example, the types of text structure or content which were related to a greater number of regressions, or longer fixations, and so on. For example, if one staged certain information high in the text, one might expect longer fixations on that information, but only when the content was familiar. One might thus be able to infer what strategies able and less able readers were using in certain text situations. Of course, if this type of data is combined with product measures, one can make more inferences and have better grounds for making them.

Dual task studies. Computers can also be used to indicate whether or not readers are expending more or less effort on specific segments of text and, thus, whether or not they are attending to the structural characteristics of the text. The method is to have readers read passages whose characteristics are clearly specified, having them expose the text sentence by sentence on a screen. At various specific points in the text, a tone is sounded to which they must respond by pressing a button. The button-press latencies are related to the processing load at the time. The approach is based on the concept of channel capacity, and it seems to yield reliable information about concentration at different points in the text (Reynolds, Standiford, & Anderson, 1978). If this can be refined somewhat, one could assess the extent to which, given a specific task, an individual is responsive to the textual indications of importance.

Self-controlled exposure procedures could also possibly be used to examine other processes in reading, such as the use of look-backs, especially when difficulties or errors are deliberately placed in the text. A further advantage of the computer is that such measurements could be taken in the context of a game situation, thus overcoming the problem of motivation difficulties normally inherent in testing procedures.

Metacognitive Measures

A considerable amount of research has recently been directed at examining what are called metacognitive skills (e.g; Baker & Brown, 1980; Brown, 1978). These are skills by which individuals monitor their own progress in tasks. The skills and strategies which readers use while reading need to be actively adjusted in the light of the state of the reader's comprehension relative to his goals. A second aspect of metacognition is the awareness of the demands of the various tasks to be engaged in and the characteristics of the individual in relation to those demands. This awareness of the strategies and when and where to use them is just as important an aspect of reading comprehension as the availability of the strategies themselves. The following assessment techniques require varying degrees of this awareness.

Self-corrections. Self-corrections (e.g., Clay, 1973) provide a potential source of information. They are closely related to miscues but are the result of readers monitoring their own reading comprehension, for instance, when a child reads a word incorrectly and then returns to correct it several words later. The use of various cue systems can be inferred from the location and delay of the self-correction and its relationship to the original error. There are two main problems with the use of self-corrections as a sole source of information. Firstly, they do not occur with high frequency under normal circumstances. Secondly, and most importantly, readers often find information that causes them to revise their earlier model, but they do not do so overtly. Nevertheless, as an adjunct and incidental source of information, self-corrections can be very informative.

Protocol analysis. Protocol analysis (e.g., Kavale & Schreiner, 1979; Olshavsky, 1976-77) is a clear example of an on-line measure which requires metacomprehension skills. This technique requires the reader not only to read aloud, but to think or process aloud, also. The reader is expected to give a fairly full account of what occurs in his head while actually reading. This differs from a retrospective account and an introspective account in that it is more on-line. Of course, one must make certain bold and questionable assumptions about the ecological validity of such a task and the time delay between reading and interjecting comments, but given these assumptions, this technique, too, provides useful information on the comprehension process. For example, it can indicate how, when, or whether an individual uses certain information and how his model changes as he progresses.

The younger the children, however, the less likely they are to be able to perform the task, since young children have less access to this metacognitive knowledge than do older children. A certain amount of training is required before children can readily use this technique, but it is especially useful as they respond to questions, since it can help detect retrieval problems or test-taking skill problems. Such training as is required is far from a waste of time. It serves to provide the child with access to his own thought processes, allowing him to monitor them himself.

Error-detection tasks. Another metacognitive task which has seen a great deal of recent experimental exposure is the use of children as critics who are asked to evaluate incomplete or faulty instructions and are supposed to say in what way the instructions are unclear or should be improved (Markman, 1979). This task requires the child to comprehend the instructions as far as they will allow, see that they are incomplete or erroneous and comment to that effect. This last demand is the big problem with this approach, since if the child does complete the task, one can say something about his comprehension and metacomprehension skills, but, if he fails to mention the mistake, one is unable to say where the failure has occurred. It is quite likely that the error was noticed but was either ascribed to his failure or otherwise explained away (Winograd & Johnston, 1980). Nonetheless, the task has its uses, especially when a teacher has a good rapport with the children. It is most useful with instructions, and of very limited use with dialogue and other similar texts. This, of course, is reasonable, since the children have a right to expect instructions to make sense but are well aware that people need not tell the truth or be consistent.

This procedure in combination with eye movement or other on-line measurement might be a very effective way of observing the strategies which readers use to deal with structural or content disruptions, which amount to simulations of natural reading comprehension problems.

Confidence rating. Different variants of this task require examinees to rate or place a probabilistic value on the correctness of their responses or multiple-choice alternatives. This approach has been debated as far back as 1929 (Greene, 1929). The principle motivations for its use have been its conceptual appeal, that it allows for the assessment of partial knowledge, and that, in combination with penalties for highly rated incorrect answers, it discourages guessing in multiple-choice tests (Echternacht, 1972).

More recently, however, interest in the metacognitive aspects of the method has increased (Anderson, Note 14). Not only can it assess the extent to which the examinee knows that he knows (or does not know), but with multiple-choice items it can also be used to assess guessing behavior if equal ratings of different alternatives are allowed (Johnston & Pearson, Note 7). The approach may also have side benefits. For example, if individuals are required to rate each alternative in a multiple-choice question, they may be forced to process every alternative, which may not normally be the case, and to process them more deeply. One could argue, on the other hand, that the extra task demands interfere with, rather than assist, the examinee's performance. This remains to be studied.

Other metacognitive tasks. Collins, Brown, Morgan, and Brewer (1977) have used underlining or circling of key words or sentences to assess children's ability to select and study main ideas. They found that children below the seventh grade and educable retarded junior high children do not generally perform such activities without being instructed to do so. When they are instructed to do so, only those children who have previously done so spontaneously seem able to underline important units. Others tend to underline unfamiliar, long, or random words. Circling key words is easier than underlining key idea units. Both, however, give evidence of the child's progress in mastering the skills of selecting important information. It might also be found useful to observe the child while he performs this task, especially if one had the child follow the words with his pencil. For example, pauses might indicate reprocessing, and lookbacks may also be evident. Note-taking is probably an even better indicator of reading comprehension. While it is conservative because of the production requirement, it nonetheless seems to give a good indication of the reader's comprehension, and hints at the processing strategies used.

Another potential assessment technique requires children to rate the relative importance of the idea units in the text. This is a readily scorable task and can provide information on the reader's understanding of the text. Brown and Smiley (1977) note that while third-, fifth-, and seventh-grade children's recall is highly related to relative importance (as rated by college students), there is differential ability across grades to explicitly rate the units. However, this may well be due to the memory demands of the task and may disappear if idea units were on cards and could be sorted into rank order of importance.

Brown and Day's (Note 4) work on summarizing skills has indicated further alternative outcome variables such as people's ability to summarize text. Even simpler components of such tasks, for example, deleting the unimportant (or the redundant) information from a passage, could give a reasonable indicator of reading comprehension.

The ultimate extension of the metacognitive approach to assessment is self-assessment. Collins, Brown, Morgan, and Brewer (1977) and Brown et al. (1980) have laid the groundwork for this, emphasizing the fact that children must be taught to monitor their own performance, by providing explicit instruction, modeling, and checklists.

Recently, this approach was put into practice by Pflaum and Pascarella (1980). They had children listen to tapes of miscues (initially someone else's, then their own) and sensitized them to the miscues which were being made. They then taught them to decide when they had made an important mistake and a less important mistake, and to take appropriate action. Thus the assessment technique became a self-assessment, and had payoffs in the children's improved performance in reading.

Summary

This section has examined a variety of aspects of assessment methodology. It has indicated some of the constraints which operate in reading comprehension assessment situations and how we might deal with them. A fair proportion of the problems can be solved by taking a broader perspective on reading comprehension, especially going beyond standardized test scores. Standardized tests have generally restricted our thinking to product-type tests when what we should be more concerned about are the comprehension processes and metacomprehension.

Concern in assessment practice has been much more over reliability than validity. It is suggested that this situation be reversed. This involves a concern not only for more natural texts and tasks, but for using the contextual variability, rather than standardizing it, to come to a better understanding of a reader's capability. More careful analysis of the demands of the tasks which we impose upon the students will provide us with greater insight into why students fail on them. Generally, it is better to assess ability using a variety of tasks as sources of information, since each provides different information. Measured reading comprehension and awareness of reading comprehension are both a function of how they are measured.

An important part of our assessment procedure should be teaching the student to use the procedure for self-assessment. This has an important instructional function but should also increase the reliability and validity of our own future assessments of the students, since they will be familiar with the tasks.

CONCLUDING SUMMARY AND IMPLICATIONS

In describing the contributions of basic research to the assessment of reading comprehension, this paper has presented a picture of what reading comprehension is and what factors affect it. It has also pointed out how these factors affect the outcome of assessments of reading comprehension and the tension between what we require of tests and what they can give us. Tuinman (1979, p. 45) claims that ". . . once test constructors included inferential comprehension items in test batteries it naturally became increasingly impossible by any technique of statistical reduction to show anything but that reading was indeed reasoning." The present paper disputes that claim somewhat by pointing out the reasons for the high correlations between reading comprehension, vocabulary, and measures of general reasoning ability. Reading involves various specific reasoning skills. It also involves the use of already possessed background knowledge, which is probably what vocabulary tests assess. It is largely the background knowledge component which affects inferencing. However, over a number of different passages in different content areas, the general knowledge required will correlate increasingly highly with measures of intelligence.

Delineating some of the relevant reasoning skills is an important task. These are the processes of reading comprehension and an important challenge for us is to come up with ways in which we might tap these using measures available to the classroom teacher and, possibly, product-type, mass-administered measures. In this concluding section we shall deal with the following topics:

1. What can and should we assess?
2. How might we assess these aspects?
3. The characteristics and interpretation of reading comprehension tests.
4. Purposes of assessment.

What Can and Should We Assess?

This paper has looked at reading comprehension in terms of a reader's systematic use of various cue systems to produce a model of the author's intended meaning. Authors may or may not provide appropriate cues, and a reader may be reading a text for a purpose other than that intended by the author. It was, hopefully, made clear that text is not a random variable which we can generalize across at will, nor is reader purpose. We must know where the information is available to the reader and what cue systems are provided in the text so that we can assess whether or not the reader is aware of them, able to use them, and capable of being independent of them. Basic research is required in order to continue to isolate the important textual and contextual variables which are involved and in order to describe their interrelationships.

In light of this, we have pointed out several cue systems which readers can use. We have also examined several cognitive reasoning strategies which readers might use to keep their comprehension abreast of their reading. There are two levels of strategies which can be described. The first set of strategies is involved with helping the reader construct a model of the meaning of the text from the information available in the text. The second set requires readers to monitor their progress toward understanding the text, detecting lapses in comprehension, and initiating strategies to rectify the difficulties.

In our assessment, then, we might consider assessing the following:

1. Awareness and use of the available cue systems in the text. The explicitness of the cue systems in the text affects how readers are able to deal with text. We should be interested in students' ability to use the cues when they are present, and their ability to deal with the text when the cues are not explicit. This "degree of explicitness" of the cues represents the extent of the assumptions which the author has made about the reader. Some potential candidates for evaluation might be:

a. Text type, e.g., narrative versus expository versus rhyming poetry. We must await and encourage further development in the area of classification of text in order to do justice to this aspect, although macrostructures such as temporal form versus hierarchical form (Thorndike, 1977) or staging (Meyer, 1975) will probably play a role in this.

b. Internal social-organizational aspects (Bruce, in press).

c. Author-reader relationships (Tierney & LaZansky, 1980).

d. Perspective.

e. Connectives.

2. We might consider directly assessing the reader's perception of the author's intentions and plans, especially when we get answers which do not match the expected ones.

3. Abilty to override cue systems which do not correspond with the reader's purposes for reading, for example, when the text is persuasive and the reader's purpose is evaluative.

4. In constructing tests we must be aware of the nature of the text, especially in terms of its relationship to the background knowledge of the reader both quantitatively and qualitatively. We should directly assess such background knowledge differences to assist us in our diagnosis of problems.

5. We might consider assessing for diagnostic purposes, whether or not dialect is causing a problem with reading.

6. Comprehending without being able to access the information later

is often of limited value. Thus, we might consider assessing retrieval strategies.

7. Production skills can be assessed separately from reading comprehension and taken into account in interpretation of production-type reading comprehension tests.
8. Information processing and use of prior knowledge should be assessed. Is the student getting information from the text? Is it stored and integrated? That is, are the various logical and pragmatic inferences being made? Inferences may be examined according to the classification system presented by Warren, Nicholas, and Trabasso (1979).
9. We might assess comprehension skills such as those suggested by Spiro (1980):
   a. schema selection;
   b. schema maintenance;
   c. schema instantiation and refinement.
10. Cognitive monitoring skills are important. For example, is the reader able to distinguish between when he does and does not comprehend? Does he have self-checking routines to evaluate this?
11. Knowledge of comprehension maintenance strategies should be assessed. For example (from Collins & Smith, 1980):
    a. ignore and read on;
    b. suspend judgement;
    c. form a tentative hypothesis;

   d. reread the current sentence(s);
   e. reread the previous context;
   f. consult an expert source.

12. From Brown and Day's (Note 4) work, one might also consider assessing knowledge of summarization rules and strategies.
13. Does the reader know when to use specific strategies, i.e., know the "triggering conditions" (Collins & Smith, 1980)? For example, given a reading purpose (e.g., reading for gist), knowing what to do when not understanding something (e.g., a single word) and the costs and benefits of strategies (e.g., skip the word and go on).
14. Assessment of children's understanding of the demands of various reading tasks would enable us to rectify such potential sources of failure.

How Might We Assess These Aspects?

All assessment of reading comprehension is indirect, in that we cannot actually see the processes or get a "pure" measure of reading comprehension alone. By understanding the basic processes involved in the assessment tasks we can understand what we do and do not know from our assessment, i.e., what inferences we can reasonably make about a child's performance. Thus we take it as "given" that we can never have a perfect measure of the cognitive processes which we are investigating. We are bound to have some error or "noise" in our assessment. However, with this in mind we will suggest some possible ways to assess the aspects which we deemed important in the previous section.

1. As the cue systems become more clearly defined, we will be able to manipulate them independently between and within texts in terms of their explicitness and appropriateness. Accurate assessment of readers′ problems with various cue systems will necessitate comparable outcome measures across different texts, thus question-generation strategies will require more complete development before this type of reader information is fully realized. However, underlining specified cues or relationships may serve as an adequate measure of recognition.

2. In assessing ability to override cue systems, one could use the same measures as used to assess awareness, but provide the reader with a different purpose for reading. Merely asking students to read for a different purpose is generally not sufficient, they must be convinced that it is in their best interests to do so. This can be done by showing the student and/or specifying the task which is to follow the reading. It can also probably be better accomplished by providing a practice task of the type to be required. A further effective alternative is to use questions embedded in the text. This technique tends to make readers adjust their purposes, and hence their allocation of attention.

3. Understanding the effects of different text variables on reader′s comprehension should show us how to select more optimally lucid texts. This will put us in an interesting dilemma as far as test construction is concerned. Should we assess reading comprehension using "perfect" passages or whould we use moderately awkward but naturally occurring passages? Do we wish to know how a reader performs given optimal conditions or suboptimal?

Perhaps it will be that we want both, since it seems unlikely that all literature will ever be optimally lucid. The contrast between performance on good versus not-so-good text for an individual may well be instructive. Perhaps we will have to explicitly teach readers how to deal with writer inconsiderateness.

4. The passage dependency of items may possibly be used as a tool to separate out the effects of prior knowledge (such as potential cultural bias) from other comprehension problems, and to examine the extent of integration of textual information. Trabasso indicates that inference questions are the most powerful assessments of comprehension; however, in order to know whether a failure to make a specific inference was due to lack of ability to do so or lack of the requisite prior knowledge, it may be useful to ask specific passage-independent, but highly relevant, prior knowledge questions. These items could not only be used to detect a lack of appropriate prior knowledge, but also qualitatively different prior knowledge which could lead to "incorrect" answers on other items. This could possibly be done most efficiently using the true-false format of Anderson and Freebody (1980). On the other hand, the classroom teacher can simply ask relevant questions and judge the extent of background knowledge prior to or subsequent to asking the inference question.

5. To answer the question "Is information which was comprehended retrievable?," one could assess comprehension immediately after reading, and reassess at a later date. Comparing performance on posttest questions when text is available versus when it is unavailable during question answering would also supply relevant information on this aspect.

6. Looking for comprehension skill deficits of the type which Spiro (1980) addresses might be done in the following manner: Schema selection would be evident by providing a clear title for a relatively vague text (strong author assumptions), and comparing performance on this text with performance on a standard text. Schema maintenance could be assessed by constructing a text which has a context-setting section, then a long section, which only really makes sense in the perspective of the first section. Errors or points of possible misinterpretation could be embedded various distances into the text. With or without embedding errors, oral reading of the text may indicate, through the type of miscues occurring, how comprehension is proceeding. Performance on these texts could be compared with performance on texts which did not make such demands.

7. To assess readers' skill at monitoring their own reading comprehension, one could build into the text systematic distortions (e.g., Markman, 1979); however, expecting a reader to subsequently voluntarily report finding the distortion is not a very reliable way of assessing error detection (Winograd & Johnston, 1980). A better way might be to combine this type of disruption (essentially a simulation of reader-generated comprehension failures) with eye movement data or self-controlled exposure times (computerized). Such distortions could also be introduced at different levels of text organization, and both aspects could be examined together. Another useful way of assessing comprehension monitoring is through the use of confidence ratings in conjunction with almost any product measure including, for example, ranking of idea units, or rating multiple choice alternatives, etc.

8. Knowledge of comprehension maintenance strategies could be assessed merely by asking readers what they do when certain situations arise, and asking them to take note of this when reading (i.e., introspection). One could even go so far as to construct text situations which would be likely to force readers to use these strategies. The think-aloud technique could be used to get information on the use of these strategies too. Getting readers to follow their reading with their finger or a pencil may give hints as to when they use certain of the strategies, and eye movement data would help detect certain strategies also.

9. Knowing when to use various strategies (i.e., recognizing the triggering conditions) and knowing the risks and payoffs of each could simply be asked directly of readers. This approach would, however, only test whether they knew the theory. To assess whether or not they actually practice similarly, one would probably need to use a think-aloud approach since this is so highly dependent on text, task, and particularly the individual, his skills, and prior knowledge. It can only really be assessed in the one-to-one situation.

10. We must endeavor to make the assessment task meaningful and interesting. Unless we do this, we risk the children's failing because of their decision not to treat this new information as "real," and not integrating it. Computers are particularly suited to this problem since there are already many commercially available games which require some reading skill to play. Increasing the reading component and assessing the player's behavior following a "clue," for example, may provide unobtrusive, ecologically valid information on the player's reading comprehension.

11. A method for assessing the effects of dialect has been demonstrated by Steffensen, Reynolds, and McClure (in press). They used the fact that speakers of Black English Vernacular (BEV) do not use inflected forms of regular verbs for past and present tenses, to show that in situations in which the inflection is the only source of temporal information, BEV speakers have more difficulty than speakers of Standard English. Their assessment instrument was a passage which the reader had to place in a time frame, yesterday, tomorrow, or anytime.

12. The extent to which probed recall is better than free recall may indicate retrieval problems rather than comprehension or storage.

Many of the above techniques require one-to-one administration. For product measures, while there are alternatives to questions, they are probably still the prime source of information. They need to be more carefully constructed in order to refine this information. In any case, a departure from the assumption that all things be kept equal in assessment and a greater concern with ways to use systematic variation in testing procedure as a source of additional information is advocated. This is in line with Vygotsky's (1978) notions of dynamic assessment which essentially concentrates on locating the conditions under which readers can or cannot perform certain tasks, rather than simply whether they can or cannot perform them. For example, the number and type of hints required to prompt the correct response to a task is a continuous and more accurate indicator of the child's level of knowledge or skill.

Question Construction

Currently, test items are not generated in a particularly meaningful way. Perhaps this can be partially rectified in light of the studies presented in this paper. For example, an amalgamation of Pearson and Johnson's (1978) and Lucas and McConkie's (1980) classification systems, along with some additional changes, may produce a diagnostic classification system. Research indicates at least two dimensions along which items should be classified. One dimension is source of the information, i.e., prior knowledge, versus the text, including location within the text. One element which must be added to this dimension relates to the structural location of the information in the text in terms of causal chains or staging, for example. The second dimension concerns the cognitive demands which the question places upon the learner. A further breakdown could also be performed on the inferential questions in this dimension on the basis of, for example, Warren, Nicholas, and Trabasso's (1979) analysis of inferences. However, it remains to be seen whether these modifications will turn out to be effective. We must begin to develop multiple choice alternatives such that by systematic production of alternatives, comparison across items will yield more information per unit of test time. This will especially be the case if the alternatives carry their own individual meaning, each representing an attractive answer given that the reader used a different specific processing strategy, a different cue system, or a (possibly culturally) different perspective. Unless we specify the task demands and goals for each question, the meaning of an incorrect answer, or even a correct answer, is unclear. Hopefully, when item development is based on meaningful skills, we will no longer select items on the basis of person discrimination (discrimination index) but rather on the basis of skill separation.

No question classification system includes items which are passage independent. Now that we know of the measurement problems caused by questions which draw on background knowledge and reasoning skills there are two opposing ways for us to respond. We could take the approach of Tuinman (1974), Royer and Cunningham (1978), and others who suggest removing items which are "biased" in these ways (Royer and Cunningham suggest that this need not be the case for predictive tests). This would involve selecting passages which dealt with generally familiar topics and eliminating items which were passage independent or involved some reasoning ability. Royer and Cunningham correctly point out that it is not possible to do either of these completely, but they contend that one should minimize these factors. It is this writer's contention that these two factors are involved in normal reading comprehension and that both are possible sources of comprehension problems.

Since we have the knowledge to know which items are tapping these areas, we should explicitly use the information which they give us rather than throw it away. Our behavior has been rather like throwing away a find of oil, having not realized that it is more than just dirty black stuff. Indeed, we should actively attempt to refine these questions so that they will give us the best possible data. To ignore these data is to pretend that reasoning and prior knowledge are not functional aspects of reading comprehension.

This seems to have been a general problem since other assessment techniques have also been rejected for what they do not tell us while what they do tell us has been ignored. Perhaps, rather than throwing it away, we should have been using this valuable information by looking at the relationships between the "flawed" measures.

Characteristics and Interpretation of Tests

Venezky (1974) claims that comprehension measurement techniques should be tied to curriculum materials. This is only partially true. Currently, it can hardly be claimed that the design of curriculum materials for teaching reading comprehension is based on a theory of cognitive functioning, and this is what we are particularly concerned about when assessing reading comprehension. Indeed, such a concern may help redirect us towards a curriculum which is more in tune with how children actually comprehend.

At all levels of assessment, an understanding of the processes which are to be assessed is crucial. It is difficult to measure something which we do not have a clear concept of. It is through such a theoretical framework that we can come to understand individual differences and how to respond to those differences which we record. This knowledge of text and text processing must be coupled with a knowledge of the demands of our assessment tasks so that we know exactly what we have asked the child to do. Thus we might gain a knowledge of the various possible sources of error.

This also allows us to minimize the number of error sources by ensuring that the child's understanding of the task is the same as ours, and that the situation is appropriate and motivating. Indeed, we should teach him directly the skills and strategies required for taking our tests.

The most important function of recent research in reading comprehension, with respect to its assessment, is largely as metaknowledge. That is, it allows us to become aware of what we do and do not know when we have assessed a child's reading comprehension, and points out what we could know. We should now be able to make improvements in both reliability and validity, but especially in validity. To date, reliability has been acceptable but validity has been somewhat less acceptable (a) in terms of the reality of the subskills under assessment, (b) because of the inadequacy of the tests to assess them effectively, and (c) because of the lack of ecological validity. We wish to de-emphasize reliability (without losing it) and emphasize validity.

Both reliability and validity must be looked at in terms of the decisions which result from our assessments. Greater understanding of strengths and weaknesses of procedures can help increase both aspects of our decision-making. Validity can be increased by better understanding of (a) what is being measured, (b) how to measure it, and (c) what factors influence performance and, hence, must be taken into account when interpreting performance. As a result of increasing the validity of the tests there should be an associated increase in reliability because of (a) clearer specification of parallel forms and (b) greater understanding of, and ability to control for, the effects of extraneous factors. These factors can potentially invalidate an instrument which, under certain circumstances, is quite valid. For example, the demands of the social situation in which assessment occurs can strongly influence the outcome of the assessment (Steffensen & Guthrie, 1980). It is this writer's opinion that informed use of measurement devices is the key to ensuring test validity, since validity can only be "built in" to tests to a certain degree. Invalidity lies as much in the claims which we make about what we have measured as in the instrument itself. In this light then, the following general cautions and considerations are presented:

1. Current reading comprehension assessment practices are unsatisfactory partly because of the constraints placed upon them. The problems stem not only from practical constraints but from conflicting theoretical concerns as well.

2. We cannot see cognitive processes. Cognitive and developmental psychology are helping solve this problem by analyzing various cognitive tasks experimentally and by showing us when the cognitive strategies are evident or at least "close to the surface," for example the think-aloud and error-detection paradigms. Teaching readers how to introspect, and what things to be aware of may also later assist us through self-report.

3. Item selection and construction can and must become more systematic. The information available from the assessment will thus become more useful.

4. Using more questions makes for greater reliability but more interitem correlation also. Unfortunately, using more items also means increased time demands and often a decrement in attention. The intent here is to increase the efficiency of the testing by making explicitly clear what we are testing and using tasks with known properties (both similar and different) in which responses provide maximum information.

5. The assessment task should be carefully analyzed and made quite explicit so that it is very clear to both assessor and assessee. Specification of the task should be such that one can understand the meaning of a "wrong" answer as well as a "right" answer. Knowing exactly what abilities we are assessing when we use a test enables us to increase the validity of the decisions and statements which we make as a result of the assessment.

6. Knowing what factors can affect our assessments can increase the reliability of our findings by allowing us to account for more of the variability in our findings and increasing the agreement on interpretation.

7. It may be that children look on reading comprehension assessment as a special task and treat any information gained through such required reading as "different" and deliberately reduce schematic integration.

8. By systematically varying the tasks which the reader must perform, we may be able to assess which aspects of the reading task may be causing problems for the reader.

9. These tasks should be highly related to normal societal demands on reading performance.

10. We must carefully consider the social aspects of our assessment both in terms of the effects of the testing situation on the perceived task, and in terms of the social import of the assessment to the individual.

11. Assessment texts should include texts which are complete enough to have macrostructures.

12. The order in which a reader reads a series of texts affects the comprehension of each. Test constructors must consider this variable in compiling tests, just as they must consider primacy and recency effects of information occurrence in each text.

13. The assumption underlying the more recent analysis of texts is that the text is a communicative device. Someone has produced it in an effort to convey meaning to another. This does not seem like a big assumption, but in reading various passages in reading comprehension tests, it becomes clear that this assumption is sometimes violated.

14. People approach different types of text differently. That is, one might not normally read an expository piece on a scientific topic in the same way as, say, a Shakespearean sonnet, or True Romances. Certain goals are more likely for certain types of text. Indeed, the motivation for reading at all (as well as for reading with different strategies) may be controlled by the type, and certainly the relevance, of the text. A youngster whose chief concern is to learn how to fix bikes may well be able to perform adequately on service manual comprehension tasks but fail on poetic or narrative forms. These should not be muddled in our assessment by giving a global test score on a nonspecific text type with a nonspecific purpose.

15. Assessment texts should be motivating or, if they are not, then this should be taken into account as a possible cause for failure when interpreting the test. Texts should include variation in those factors which we think may affect the individual's performance, such as dialect and background knowledge. These factors should also be directly assessed through the questions.

16. It is possible to have situations in which a reader may have used all the appropriate strategies, but have developed an inappropriate model of the author's meaning because of a deficient or different background knowledge. We should begin to take into account in our assessments the possibility of alternative plausible interpretations of text. This might be done through the use of free recalls, or by allowing comments on multiple choice questions, which could indicate to an assessor the reasoning underlying an answer.

17. A child's readiness to respond to a question is determined by a number of variables besides his understanding of the question.

18. Probably the best way to assess reading comprehension is to use two or three different types of measure, each based on different assumptions and with different sources of error. This gives firmer grounds for such inferences as one wishes to make and allows a greater variety of inferences.

We shall now consider the implications in terms of the distinction made earlier between the three levels of decision-making: administrative, diagnostic, and selection.

Administrative Assessment

From the earlier discussion of text and task variables, it seems clear that accounting for all of them in a single test which also requires a range of difficulty and passages of reasonable length, and at the same time has a fairly strict time limit, would be difficult to say the least. However, since the decisions are not at the level of the individual pupil (or test paper), is not necessary for all children to take the same test. The best way to deal with this is to devise a test using all the parameters, then use multiple matrix sampling techniques to produce a series of separate test booklets containing parts of the test. Thus, several pupils take the test between them. Selection of items on different forms can be made in a systematic way so that each child gets items with certain relationships between them, and certain items which other children also get. Thus, very sensitive program evaluation could be done, and more meaningful decisions could be made. One would have, in a sense, program diagnostic information.

Diagnostic Assessment

The major thrust must be towards individual, diagnostic assessment in which the aim is to specify not merely that the child can or cannot perform a particular task, but to specify the conditions under which the child can or cannot do it. In order to assess in this manner, the assessor really needs to have a theory with which to generate testable hypotheses about the reasons for the reader's performance. He also needs to know the capabilities and limitations of the various assessment techniques. Thirdly, he must be well aware of the factors which represent the "conditions under which" comprehension occurred. Awareness of each of these factors can make the teacher a more sensitive observer of behavior, and thus, ultimately, a continuous, unobtrusive assessor. At that point, instruction and assessment effectively become one. The assessment is a matter of locating the "leading edge" of a child's performance and defining the generality of his performance. The assessment requires presenting the child with tasks in the range of optimal challenge and examining how much and what type of extra prompting or other assistance is needed to enable him to deal successfully with the task.

It is also concluded that if our assessment procedures are reasonable, then they should be taught to the child, since not only will that reduce the extraneous information which we get in later assessments, but it will allow the child to monitor his own reading comprehension. This type of self-monitoring instruction is a must. We should make completely clear the assessment demands of our instruction and give the child practice at using the techniques.

We should recognize the importance in our assessment procedures of such variables as retrieval strategies, and teach them. Rather than being concerned about them as potentially confounding by-products, we would do well to look more carefully at them and their role in reading comprehension. Perhaps what we require is a systematic way of selecting assessment procedures for given situations, readers, tasks, and texts.

Selection Assessment

As explained previously, this type of test generally needs to be group administered, but the decisions are to be made at the individual level. This presents an awkward trade-off situation in that we need much information which we cannot get because we are forced to use the time inefficiently. Since the test must cater to a range of ability, each pupil spends a good part of his test time on materials that are inappropriate and tell us little. While we can make the assessment somewhat more efficient by using systematic item and alternative selection, we still cannot eliminate the problem.

We contend that these tests may be used for initial grouping practices in the classrooms and schools, but only with great caution. If they are used with a good understanding of their problems and inaccuracies, they may be functional as initial gross screening devices. However, we believe that currently too much emphasis is placed on the results of such tests (for example, Title I placement) and that teachers should be far more wary of the scores and more accepting of their own analyses. It is hoped that the training of teachers will begin to include more comprehensive work on the skills teachers require for the dynamic interactive type of assessment discussed earlier.

Computers may provide a partial solution to certain problems associated with motivation, efficiency, and time constraints. By using branching programs which adjust text, item type, and difficulty on the basis of initial item performance, more efficiency could be achieved. Thus, some of the benefits of individualized assessment could be gained without additional time investment from the teacher. The motivational aspect may be dealt with by building these into computer games which pupils would play for amusement and which would incidentally require certain reading skills. Assessment could be continually adjusted, assessing comprehension via the child's responses such as the number of "hints" required (each costing a quantity of treasure) to gain certain crucial information from a segment of text. As the initial investment required for the necessary equipment becomes increasingly less expensive, such developments will certainly prove very useful.

In conclusion, it is hoped that this paper has provided what amounts to assessment metaknowledge such that its readers, having assessed students' reading comprehension, have a better idea of what the students can and cannot do and how they might go about finding out more.

Reference Notes

1. Derrick, D. *Three aspects of reading comprehension as measured by tests of different lengths* (Research Bulletin 53-8). Princeton, N.J.: Educational Testing Service, 1953.
2. Pace, A. J. *The influence of world knowledge in children's comprehension of short narrative passages*. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.
3. Wilensky, R. *Understanding goal-based stories* (Research Report No. 140). New Haven, Conn.: Yale University, Department of Computer Science, 1978.
4. Brown, A. L., & Day, J. D. *Developmental trends in the use of summarization rules*. Paper presented at the meeting of the American Educational Research Association, Boston, 1980.
5. Stein, N. L., & Glenn, C. G. *The role of structural variation in children's recall of simple stories*. Paper presented at the meeting of the Society for Research in Child Development, New Orleans, March 1977.
6. Adams, M., Bruce, B., Cohen, P., Collins, A., Gentner, D., Rubin, A., Smith, E., Starr, B., Starr, K., & Steinberg, C. *Progress Report 3 on the BBN Group of the Center for the Study of Reading* (BBN Report No. 4461). Cambridge, Mass.: Bolt Beranek and Newman, July 1980.

7. Johnston, P. H., & Pearson, P. D. *Explicitness of connectives and content familiarity as determinants of reading comprehension*. Paper presented at the annual meeting of the National Reading Conference, San Diego, December 1980.

8. Raphael, T. E., Freebody, P., Fritz, M., Myers, A. C., & Tirre, W. C. *Contrasting the effects of some text variables on comprehension and ratings of comprehensibility*. Paper presented at the annual meeting of the American Educational Research Association, Boston 1980.

9. Harste, J., & Burke, C. *Reexamining retellings as comprehension devices*. Paper presented at the National Reading Conference, San Antonio, December 1979.

10. Beck, I. L., McKeown, M. G., McCaslin, E. S., & Burkes, A. M. *Instructional dimensions that may affect reading comprehension: Examples from two commercial programs* (Tech. Rep.). Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1979.

11. Brown, A. L., & Ferrara, R. A. *Diagnosing zones of potential development: An alternative to standardized testing*? (Tech. Rep. to appear). Urbana: University of Illinois, Center for the Study of Reading, in preparation.

12. Frederiksen, C. H. *Inference structure and the structure of children's discourse*. Paper presented at the Symposium on the Development of Discourse Processing Skills, Society for Research in Child Development, New Orleans, 1977.

13. Turner, A., & Greene, E. *The construction and use of a propositional text base*. Report from the Institute for the Study of Intellectual Behavior, University of Colorado, April 1977.

14. Anderson, R. I. *Confidence testing: A means to assess a metacognitive component of test item responses*. Unpublished manuscript, University of Illinois, 1980.

References

Adams, M., & Bruce, B. Background knowledge and reading comprehension (Reading Education Rep. No. 13). Urbana: University of Illinois, Center for the Study of Reading, January 1980. (ERIC Document Reproduction Service No. ED 181 431)

Adams, M. J., & Collins, A. A schema-theoretic view of reading comprehension (Tech. Rep. No. 32). Urbana: University of Illinois, Center for the Study of Reading, April 1977. (ERIC Document Reproduction Service No. ED 142 971)

Aiken, E. G., Thomas, G. S., & Shennum, W. A. Memory for a lecture: Effects of notes, lecture rate and informational density. Journal of Educational Psychology, 1975, 67, 439-444.

Anderson, R. C. Schema-directed processes in language comprehension (Tech. Rep. No. 50). Urbana: University of Illinois, Center for the Study of Reading, July 1977. (ERIC Document Reproduction Service No. ED 142 977)

Anderson, R. C., & Freebody, P. Vocabulary knowledge (Tech. Rep. No. 136). Urbana: University of Illinois, Center for the Study of Reading, August 1979. (ERIC Document Reproduction Service No. ED 177 480)

Anderson, R. C., & Pichert, J. W. Recall of previously unrecallable information following a shift in perspective (Tech. Rep. No. 41). Urbana: University of Illinois, Center for the Study of Reading, April 1977. (ERIC Document Reproduction Service No. ED 142 974)

Anderson, R. C., Spiro, R. J., & Anderson, M. C. Schemata as scaffolding for the representation of information in connected discourse (Tech. Rep. No. 24). Urbana: University of Illinois, Center for the Study of Reading, March 1977. (ERIC Document Reproduction Service No. ED 136 236)

Anderson, T. H., Wardrop, J. L., Hively, W., Muller, K. E., Anderson, R. I., Hastings, C. N., & Fredericksen, J. Development and trial of a model for developing domain referenced tests of reading comprehension (Tech. Rep. No. 86). Urbana: University of Illinois, Center for the Study of Reading, May 1978. (ERIC Document Reproduction Service No. ED 157 036)

Andrich, D., & Godfrey, J. R. Hierarchies in the skills of Davis' Reading Comprehension Test Form D: An empirical investigation using a latent trait model. Reading Research Quarterly, 1978, 14, 183-200.

Asher, S. R. Influence of topic interest on black and white children's reading comprehension (Tech. Rep. No. 99). Urbana: University of Illinois, Center for the Study of Reading, July 1978. (ERIC Document Reproduction Service No. ED 159 661)

Baker, L. Comprehension monitoring: Identifying and coping with text confusions (Tech. Rep. No. 145). Urbana: University of Illinois, Center for the Study of Reading, September 1979. (ERIC Document Reproduction Service No. ED 177 525)

Baker, L., & Brown, A. L. Metacognitive skills and reading (Tech. Rep. No. 188). Urbana: University of Illinois, Center for the Study of Reading, November 1980.

Baker, L., & Stein, N. L. The development of prose comprehension skills (Tech. Rep. No. 102). Urbana: University of Illinois, Center for the Study of Reading, September 1978. (ERIC Document Reproduction Service No. ED 159 663)

Bateman, D., Frandsen, K., & Dedmon, D. Dimensions of "Lecture Comprehension": A factor analysis of listening test items. Journal of Communication, 1964, 14, 183-189.

Bormuth, J. R. On the theory of achievement test items. Chicago: University of Chicago Press, 1970.

Bower, G. H. Experiments on story comprehension and recall. Discourse Processes, 1978, 1, 211-231.

Bransford, J. D. Human cognition: Learning, understanding and remembering. Belmont, Cal.: Wadsworth, 1979.

Bransford, J. D., & Johnson, M. K. Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 717-726.

Bransford, J. D., & Johnson, M. K. Consideration of some problems of comprehension. In W. G. Chase (Ed.), Visual information processing. New York: Academic Press, 1973.

Brown, A. L. The development of memory: Knowing, knowing about knowing, and knowing how to know. In H. W. Reese (Ed.), Advances in child development and behavior (Vol. 10). New York: Academic Press, 1975.

Brown, A. L. Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), Advances in instructional psychology. Hillsdale, N.J.: Erlbaum, 1978.

Brown, A. L., & Campione, J. C. Inducing flexible thinking: The problem of access (Tech. Rep. No. 156). Urbana: University of Illinois, Center for the Study of Reading, January 1980. (ERIC Document Reproduction Service No. ED 181 428)

Brown, A. L., Campione, J. C., & Day, J. Learning to learn (Tech. Rep. No. 189). Urbana: University of Illinois, Center for the Study of Reading, November 1980.

Brown, A. L., & French, L. A. Construction and reconstruction of logical sequences using causes or consequences as the point of departure. Child Development, 1976, 47, 930-940.

Brown, A. L., & Murphy, M. D. Reconstruction of arbitrary versus logical sequences by preschool children. Journal of Experimental Child Psychology, 1975, 20, 307-326.

Brown, A. L., & Smiley, S. S. The development of strategies for studying prose passages (Tech. Rep. No. 66). Urbana: University of Illinois, Center for the Study of Reading, October 1977. (ERIC Document Reproduction Service No. ED 145 371)

Bruce, B. C. A social interaction model of reading (Tech. Rep. to appear). Urbana: University of Illinois, Center for the Study of Reading, in press.

Canney, G., & Winograd, P. Schemata for reading and reading comprehension performance (Tech. Rep. No. 120). Urbana: University of Illinois, Center for the Study of Reading, April 1979. (ERIC Document Reproduction Service No. ED 169 520)

Carroll, J. B. Learning from verbal discourse in educational media: A review of the literature, 1971. (ERIC Document Reproduction Service No. ED 058 771)

Chi, M. T. H. Short-term memory limitations in children: Capacity or processing deficits? Memory and Cognition, 1976, 4, 559-572.

Chi, M. T. H. Knowledge development and memory performance. To appear in M. Friedman, J. P. Das, & N. O'Connor (Eds.), Intelligence and learning. New York: Plenum Press, in press.

Clay, M. M. Reading: The patterning of complex behavior. Auckland, New Zealand: Heinemann Educational Books, 1973.

Clay, M. M. Early childhood and cultural diversity. Reading Teacher, 1976, 29, 333-342.

Clements, P. The effects of staging on recall from prose. Unpublished doctoral dissertation, Cornell University, 1976.

Coleman, E. B. Developing a technology of written instruction: Some determinants of the complexity of prose. In E. Rothkopf & P. Johnson (Eds.), Verbal learning research and the technology of written instruction. New York: Columbia University Teachers College Press, 1971.

Collins, A., Brown, J. S., & Larkin, K. M. Inference in text understanding (Tech. Rep. No. 40). Urbana: University of Illinois, Center for the Study of Reading, December 1977. (ERIC Document Reproduction Service No. ED 150 547)

Collins, A., Brown, A. L., Morgan, J. L., & Brewer, W. F. The analysis of reading tasks and texts (Tech. Rep. No. 43). Urbana: University of Illinois, Center for the Study of Reading, April 1977. (ERIC Document Reproduction Service No. ED 145 404)

Collins, A., & Smith, E. Teaching the process of reading comprehension (Tech. Rep. No. 182). Urbana: University of Illinois, Center for the Study of Reading, September 1980.

Danner, F. W. Children's understanding of intersentence organization in the recall of short descriptive passages. Journal of Educational Psychology, 1976, 68, 174-183.

Davis, F. B. Fundamental factors of comprehension in reading. Psychometrika, 1944, 9, 185-197.

Davis, F. B. Research in comprehension in reading. Reading Research Quarterly, 1968, 3, 499-545.

Davis, F. B. Psychometric research on comprehension in reading. Reading Research Quarterly, 1972, 7, 628-678.

Dooling, D. J., & Lachman, R. Effects of comprehenson on retention of prose. Journal of Experimental Psychology, 1971, 88, 216-222.

Drahozal, E. C., & Hanna, G. S. Reading comprehension subscores: Pretty bottles for ordinary wine. Journal of Reading, 1978, 21, 416-420.

Echternacht, G. J. The use of confidence testing in objective tests. Review of Educational Research, 1972, 42, 217-236.

Farr, R., & Tuinman, J. J. The dependent variable: Measurement issues in reading research. Reading Research Quarterly, 1972, 7, 413-423.

Flavell, J. H., & Wellman, H. M. Metamemory. In R. V. Kail & J. W. Hagen (Eds.), Perspectives on the development of memory and cognition. Hillsdale, N.J.: Erlbaum, 1977.

Frederiksen, C. H. Acquisition of semantic information from discourse: Effects of repeated exposures. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 158-169.

Freebody, P. Effects of vocabulary difficulty and text characteristics on children's reading comprehension. Unpublished doctoral dissertation, University of Illinois, September 1980.

Gearhart, M., & Hall, W. S. Internal state words: Cultural and situational variation in vocabulary usage (Tech. Rep. No. 115). Urbana: University of Illinois, Center for the Study of Reading, February 1979. (ERIC Document Reproduction Service No. ED 165 131)

Goetz, E. T., Anderson, R. C., & Schallert, D. L. The representation of sentences in memory (Tech. Rep. No. 144). Urbana: University of Illinois, Center for the Study of Reading, September 1979. (ERIC Document Reproduction Service No. ED 177 527)

Goodman, K. S. *The psycholinguistic nature of the reading process*. Detroit: Wayne State University Press, 1968.

Green, G. M. *Organization, goals, and comprehensibility in narratives: Newswriting, a case study* (Tech. Rep. No. 132). Urbana: University of Illinois, Center for the Study of Reading, July 1979. (ERIC Document Reproduction Service No. ED 174 949)

Green, G. M., & Laff, M. O. *Five-year-olds' recognition of authorship by literary style* (Tech. Rep. No. 181). Urbana: University of Illinois, Center for the Study of Reading, September 1980.

Greene, E. B. Achievement and confidence on true-false tests of college students. *Journal of Abnormal and Social Psychology*, 1929, *23*, 467-478.

Grice, H. P. Logic and conversation. In P. Cole and J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3): *Speech acts*. New York: Academic Press, 1975.

Grimes, J. *The thread of discourse*. The Hague: Mouton, 1975.

Guthrie, J. T. Models of reading disability. *Journal of Educational Psychology*, 1973, *65*, 9-18.

Hall, W. S. *Testing for competence: Changing the criterion* (Working Paper No. 4, Part 1). New York: Rockefeller University Laboratory of Comparative Human Cognition and the Institute for Comparative Human Development, 1977.

Hall, W. S., Reder, S., & Cole, M. Story recall in young black and white children: Effects of racial group membership, race of experimenter, and dialect. Developmental Psychology, 1975, 11, 828-834.

Hall, W. S., & Tirre, W. C. The communicative environment of young children: Social class, ethnic, and situational differences (Tech. Rep. No. 125). Urbana: University of Illinois, Center for the Study of Reading, May 1979. (ERIC Document Reproduction Service No. ED 170 788)

Hanna, G. S., & Oaster, T. R. Toward a unified theory of context dependence. Reading Research Quarterly, 1978, 14, 226-243.

Johnson, R. E. Abstractive processes in the remembering of prose. Journal of Educational Psychology, 1974, 66, 772-779.

Kantor, R. N. Anomaly, inconsiderateness, and linguistic competence. In D. Gulstad (Ed.), Papers from the 1977 Mid-America Linguistics Conference. Columbia: University of Missouri, 1978.

Kavale, K., & Schreiner, R. The reading processes of above average and average readers: A comparison of the use of reasoning strategies in responding to standardized comprehension measures. Reading Research Quarterly, 1979, 15, 102-128.

Keogh, B. K., & Margolis, J. Learn to labor and to wait: Attention problems of children with learning disorders. Journal of Learning Disabilities, 1976, 9, 276-282.

Kintsch, W., & Keenan, J. Reading rate and retention as a function of the number of propositions on the base structure of sentences. _Cognitive Psychology_, 1973, _5_, 257-274.

Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G., & Keenan, J. M. Comprehension and recall of text as a function of content variables. _Journal of Verbal Learning and Verbal Behavior_, 1975, _14_, 196-214.

Kintsch, W., & van Dijk, T. A. Toward a model of text comprehension and production. _Psychological Review_, 1978, _85_, 363-394.

Labov, W. _Language in the inner city: Studies in the Black English Vernacular_. Philadelphia: University of Pennsylvania Press, 1972.

Lahey, B. B., McNees, M. P., & Brown, C. C. Modification of deficits in reading for comprehension. _Journal of Applied Behavior Analysis_, 1973, _6_, 475-480.

Lucas, P. A., & McConkie, G. W. The definition of test items: A descriptive approach. _American Educational Research Journal_, 1980, _17_, 133-140.

Lumsden, J. Test theory. _Annual Review of Psychology_, 1976, _27_, 251-280.

Mandler, J. M., & Johnson, N. S. Remembrance of things parsed: Story structure and recall. _Cognitive Psychology_, 1977, _9_, 111-151.

Markman, E. M. Realizing that you don't understand: Elementary school children's awareness of inconsistencies. _Child Development_, 1979, _50_, 643-655.

Marshall, N., & Glock, M. D. Comprehension of connected discourse: A study into the relationships between the structure of text and information recalled. Reading Research Quarterly, 1978, 14, 10-56.

McConkie, G. Learning from text. Review of Research in Education, 1978, 5, 3-48.

McConkie, G. W., Hogaboam, T. W., Wolverton, G. S., Zola, D., & Lucas, P. A. Toward the use of eye movements in the study of language processing (Tech. Rep. No. 134). Urbana: University of Illinois, Center for the Study of Reading, August 1979. (ERIC Document Reproduction Service No. ED 174 968)

McConkie, G. W., Rayner, K., & Wilson, S. J. Experimental manipulation of reading strategies. Journal of Educational Psychology, 1973, 65, 1-8.

Meyer, B. J. F. The organization of prose and its effect on memory. The Hague: Moulton, 1975.

Millman, J., Bishop, C. H., & Ebel, R. An analysis of test-wiseness. Educational and Psychological Measurement, 1965, 25, 707-726.

Minsky, M. A framework for representing knowledge. In P. H. Winston (Ed.), The psychology of computer vision. New York: McGraw-Hill, 1975.

Myers, M., & Paris, S. Children's metacognitive knowledge about reading. Journal of Educational Psychology, 1978, 70, 680-690.

Newsome, R. S., & Gaite, J. H. Prose learning: Effects of pretesting and reduction of passage length. Psychological Reports, 1971, 28, 123-129.

Nicholls, J. G. Quality and equality in intellectual development: The role of motivation in education. American Psychologist, 1979, 34, 1071-1084.

Norman, D. A., & Bobrow, D. G. On data-limited and resource limited processes. Cognitive Psychology, 1975, 7, 44-66.

Olshavsky, J. E. Reading as problem solving: An investigation of strategies. Reading Research Quarterly, 1977, 12, 654-674.

Omanson, R. C. The narrative analysis. Unpublished doctoral dissertation, University of Minnesota, 1979.

Omanson, R. C., Warren, W. H., Trabasso, T. Goals, inferential comprehension and recall of stories by children. Discourse Properties, 1978, 1, 337-354.

Ortony, A. Some psycholinguistic aspects of metaphor (Tech. Rep. No. 112). Urbana: University of Illinois, Center for the Study of Reading, January 1979. (ERIC Document Reproduction Service No. ED 165 115)

Otto, W., Barrett, J. C., & Koenke, K. Assessment of children's statements of the main idea in reading. Proceedings of the International Reading Association, 1969, 13, 692-697.

Paivio, A. Imagery and verbal processes. New York: Holt, Rinehart & Winston, 1971.

Paris, S. C., & Upton, L. R. Children's memory for inferential relationships in prose. Child Development, 1976, 47, 660-668.

Pearson, P. D., & Johnson, D. D. Teaching reading comprehension. New York: Holt, Rinehart & Winston, 1978.

Pellegrino, J. W., & Glaser, R. Components of inductive reasoning. In R. Snow, P. A. Fedinco, & W. Montague (Eds.), Aptitude, learning and instruction: Cognitive process analysis. Hillsdale, N.J.: Erlbaum, in press.

Pflaum, S. W., & Pascarella, E. T. Interactive effects of prior reading achievement and training in context on the reading of learning disabled children. Reading Research Quarterly, 1980, 16, 138-158.

Pugh, R. C., & Brunza, J. J. Effects of a confidence weighted scoring system on measures of test reliability and validity. Educational and Psychological Measurement, 1975, 35, 73-78.

Pyrczak, F. A responsive note on measures of the passage dependence of reading comprehension test items. Reading Research Quarterly, 1975-76, 11, 112-117.

Raphael, T. E., Winograd, P., & Pearson, P. D. Strategies children use in answering questions. Twenty-Ninth Yearbook of National Reading Conference, in press.

Reynolds, R. E., & Anderson, R. C. Influence of questions on the allocation of attention during reading (Tech. Rep. No. 183). Urbana: University of Illinois, Center for the Study of Reading, October 1980.

Reynolds, R. E., Standiford, S. N., & Anderson, R. C. Distribution of reading time when questions are asked about a restricted category of text (Tech. Rep. No. 83). Urbana: University of Illinois, Center for the Study of Reading, April 1978. (ERIC Document Reproduction Service No. ED 153 206)

Royer, J. M., & Cable, G. W. Facilitated learning in connected discourse. *Journal of Educational Psychology*, 1975, *67*, 116-123.

Royer, J. M., & Cunningham, D. J. *On the theory and measurement of reading comprehension* (Tech. Rep. No. 91). Urbana: University of Illinois, Center for the Study of Reading, June 1978. (ERIC Document Reproduction Service No. ED 157 040)

Rumelhart, D. E. Understanding and summarizing brief stories. In D. LaBerge & J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension*. Hillsdale, N.J.: Erlbaum, 1977.

Rumelhart, D. E., & Ortony, A. The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Erlbaum, 1977.

Rystrom, R. Toward defining comprehension: A second report. *Journal of Reading Behavior*, 1970, *2*, 144-157.

Sarnacki, R. E. Examination of test wiseness in the cognitive test domain. *Review of Educational Research*, 1979, *49*, 252-279.

Seigler, R. S. Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development*, in press.

Schank, R. C. The structure of episodes in memory. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.

Schank, R. C., & Abelson, R. P. Scripts, plans, goals and understanding. Hillsdale, N.J.: Erlbaum, 1977.

Slakter, M. J., Koehler, R. A., & Hampton, S. H. Grade level, sex, and selected aspects of test-wiseness. Journal of Educational Measurement, 1970, 7, 247-254.

Spearitt, D. Identification of subskills of reading comprehension by maximum likelihood factor analysis. Reading Research Quarterly, 1972, 8, 92-111.

Spielberger, C. D., Anton, W. D., & Bedell, J. The nature and treatment of test anxiety. In M. Zuckerman & C. D. Spielberger (Eds.), Emotions and anxiety: New concepts, methods and applications. Hillsdale, N.J.: Erlbaum, 1976.

Spiro, R. J. Inferential reconstruction in memory for connected discourse (Tech. Rep. No. 2). Urbana: University of Illinois, Center for the Study of Reading, October 1975. (ERIC Document Reproduction Service No. ED 136 187)

Spiro, R. J. Schema theory and reading comprehension: New directions (Tech. Rep. No. 191). Urbana: University of Illinois, Center for the Study of Reading, December 1980.

Steffensen, M. S., & Guthrie, L. F. Effects of situation on the verbalization of black inner-city children (Tech. Rep. No. 180). Urbana: University of Illinois, Center for the Study of Reading, September 1980.

Steffensen, M. S., Reynolds, R. C., & McClure, E. Black English Vernacular and reading comprehension: A cloze study of third, sixth, and ninth graders (Tech. Rep. No. 199). Urbana: University of Illinois, Center for the Study of Reading, February 1981.

Stein, N. L., & Glenn, C. G. An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), Discourse processing: Multidisciplinary perspective. Hillsdale, N.J.: Ablex, 1978.

Stein, N. L., & Nezworski, T. The effects of organization and instructional set on story memory (Tech. Rep. No. 68). Urbana: University of Illinois, Center for the Study of Reading, January 1978. (ERIC Document Reproduction Service No. ED 149 327)

Sternberg, R. J. Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities. Hillsdale, N.J.: Erlbaum, 1977.

Thorndike, E. L. Reading as reasoning: A study of mistakes in paragraph reading. Journal of Educational Research, 1917, 8, 323-332.

Thorndike, P. Cognitive structures in comprehension and memory of narrative discourse. Cognitive Psychology, 1977, 9, 77-110.

Thorndike, R. L. Reading as reasoning. Reading Research Quarterly, 1974, 9, 137-147.

Tierney, R. J., & LaZansky, J. The rights and responsibilities of readers and writers: A contractual agreement (Reading Education Rep. No. 15). Urbana: University of Illinois, Center for the Study of Reading, January 1980. (ERIC Document Reproduction Service No. ED 181 447)

Trabasso, T. *On the making of inferences during reading and their assessment* (Tech. Rep. No. 157). Urbana: University of Illinois, Center for the Study of Reading, January 1980. (ERIC Document Reproduction Service No. ED 181 429)

Tuinman, J. J. Determining the passage-dependency of comprehension questions in 5 major tests. *Reading Research Quarterly*, 1974, *9*, 207-223.

Tuinman, J. J. Reading is recognition--When reading is not reasoning. In J. C. Harste & R. R. Carey (Eds.), *New perspectives on comprehension* (Monograph in Language and Reading Studies No. 3). Bloomington: Indiana University, 1979. Pp. 38-48.

Tulving, E. Relationship between encoding specificity and levels of processing. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing and human memory*. Hillsdale, N.J.: Erlbaum, 1978.

van Dijk, T. A. Semantic macro-structures and knowledge frames in discourse comprehension. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, N.J.: Erlbaum, 1977.

Venezky, R. L. *Testing in reading: Assessment and instructional decision-making*. National Council of Teachers of English and ERIC Clearinghouse on Reading and Communication Skills, 1974.

Vernon, P. E. The determinants of reading comprehension. *Educational and Psychological Measurement*, 1962, *22*, 269-286.

Vygotsky, L. S. Mind in society: The development of higher psychological processes (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. and trans.). Cambridge, Mass.: Harvard University Press, 1978.

Wahlstrom, M., & Boersma, F. J. The influence of test-wiseness upon achievement. Educational and Psychological Measurement, 1968, 28, 413-420.

Warren, W. H., Nicholas, D. N., & Trabasso, T. Event chains and inferences in understanding narratives. In R. Freedle (Ed.), New directions in discourse processing: Advances in discourse processes (Vol. 2). Hillsdale, N.J.: Erlbaum, 1979.

Washington, E. D. The classification of reading comprehension test items and its relation to the performance of selected racial groups. Unpublished doctoral dissertation, University of Iowa, July 1979.

Weiner, B. Theories of motivation: From mechanism to cognition. Chicago: Markhan, 1972.

Winograd, P., & Johnston, P. Comprehension monitoring and the error detection paradigm (Tech. Rep. No. 153). Urbana: University of Illinois, Center for the Study of Reading, January 1980. (ERIC Document Reproduction Service No. ED 181 425)

Yendovitskaya, T. V. Development of memory. In A. V. Saparozhets & D. B. Elkonin (Eds.), The psychology of preschool children. Cambridge, Mass.: M.I.T. Press, 1971.

CENTER FOR THE STUDY OF READING

READING EDUCATION REPORTS

Adams, M. J., Anderson, R. C., & Durkin, D. *Beginning Reading: Theory and Practice* (No. 3), November 1977. (ERIC Document Reproduction Service No. ED 151 722, 15p., PC-$2.00, MF-$.91)

Adams, M., & Bruce, B. *Background Knowledge and Reading Comprehension* (No. 13), January 1980. (ERIC Document Reproduction Service No. ED 181 431, 48p., PC-$3.65, MF-$.91)

Anderson, R. C., & Freebody, P. *Vocabulary Knowledge and Reading* (No. 11), August 1979. (ERIC Document Reproduction Service No. ED 177 470, 52p., PC-$5.30, MF-$.91)

Anderson, T. H. *Another Look at the Self-Questioning Study Technique* (No. 6), September 1978. (ERIC Document Reproduction Service No. ED 163 441, 19p., PC-$2.00, MF-$.91)

Anderson, T. H., Armbruster, B. B., & Kantor, R. N. *How Clearly Written are Children's Textbooks? Or, Of Bladderworts and Alfa* (includes a response by M. Kane, Senior Editor, Ginn and Company) (No. 16), August 1980. (ERIC Document Reproduction Service No. ED 192 275, 63p., PC-$5.30, MF-$.91)

Asher, S. R. *Sex Differences in Reading Achievement* (No. 2), October 1977. (ERIC Document Reproduction Service No. ED 146 567, 30p., PC-$3.65, MF-$.91)

Baker, L. *Do I Understand or Do I not Understand: That is the Question* (No. 10), July 1979. (ERIC Document Reproduction Service No. ED 174 948, 27p., PC-$3.65, MF-$.91)

Bruce, B. *What Makes a Good Story?* (No. 5), June 1978. (ERIC Document Reproduction Service No. ED 158 222, 16p., PC-$2.00, MF-$.91)

Collins, A., & Haviland, S. E. *Children's Reading Problems* (No. 8), June 1979. (ERIC Document Reproduction Service No. ED 172 188, 19p., PC-$2.00, MF-$.91)

Durkin, D. *Comprehension Instruction--Where are You?* (No. 1), October 1977. (ERIC Document Reproduction Service No. ED 146 566, 14p., PC-$2.00, MF-$.91)

Durkin, D. *What is the Value of the New Interest in Reading Comprehension?* (No. 19), November 1980.

Jenkins, J. R., & Pany, D. *Teaching Reading Comprehension in the Middle Grades* (No. 4), January 1978. (ERIC Document Reproduction Service No. ED 151 756, 36p., PC-$3.65, MF-$.91)

Joag-dev, C., & Steffensen, M. S. *Studies of the Bicultural Reader: Implications for Teachers and Librarians* (No. 12), January 1980. (ERIC Document Reproduction Service No. ED 181 430, 28p., PC-$3.65, MF-$.91)

Pearson, P. D., & Kamil, M. L. *Basic Processes and Instructional Practices in Teaching Reading* (No. 7), December 1978. (ERIC Document Reproduction Service No. ED 165 118, 29p., PC-$3.65, MF-$.91)

Rubin, A. *Making Stories, Making Sense* (includes a response by T. Raphael and J. LaZansky) (No. 14), January 1980. (ERIC Document Reproduction Service No. ED 181 432, 42p., PC-$3.65, MF-$.91)

Schallert, D. L., & Kleiman, G. M. *Some Reasons Why Teachers are Easier to Understand than Textbooks* (No. 9), June 1979. (ERIC Document Reproduction Service No. ED 172 189, 17p., PC-$2.00, MF-$.91)

Steinberg, C., & Bruce, B. *Higher-Level Features in Children's Stories: Rhetorical Structure and Conflict* (No. 18), October 1980.

Taylor, M., & Ortony, A. *Figurative Devices in Black Language: Some Socio-Psycholinguistic Observations* (No. 20), May 1981.

Tierney, R. J., & LaZansky, J. *The Rights and Responsibilities of Readers and Writers: A Contractual Agreement* (includes responses by R. N. Kantor and B. B. Armbruster) (No. 15), January 1980. (ERIC Document Reproduction Service No. ED 181 447, 32p., PC-$3.65, MF-$.91)

Tierney, R. J., Mosenthal, J., & Kantor, R. N. *Some Classroom Applications of Text Analysis: Toward Improving Text Selection and Use* (No. 17), August 1980. (ERIC Document Reproduction Service No. ED 192 251, 43p., PC-$3.65, MF-$.91)

CENTER FOR THE STUDY OF READING

TECHNICAL REPORTS

Adams, M. J. Failures to Comprehend and Levels of Processing in Reading (No. 37), April 1977. (ERIC Document Reproduction Service No. ED 145 410, 51p., PC-$5.30, MF-$.91)

Adams, M. J. Models of Word Recognition (No. 107), October 1978. (ERIC Document Reproduction Service No. ED 163 431, 93p., PC-$6.95, MF-$.91)

Adams, M. J. What Good is Orthographic Redundancy? (No. 192), December 1980.

Adams, M. J., & Collins, A. A Schema-Theoretic View of Reading Comprehension (No. 32), April 1977. (ERIC Document Reproduction Service No. ED 142 971, 49p., PC-$3.65, MF-$.91)

Alessi, S. M., Anderson, T. H., & Biddle, W. B. Hardware and Software Considerations in Computer Based Course Management (No. 4), November 1975. (ERIC Document Reproduction Service No. ED 134 928, 21p., PC-$2.00, MF-$.91)

Alessi, S. M., Anderson, T. H., & Goetz, E. T. An Investigation of Lookbacks During Studying (No. 140), September 1979. (ERIC Document Reproduction Service No. ED 177 494, 40p., PC-$3.65, MF-$.91)

Anderson, R. C. Schema-Directed Processes in Language Comprehension (No. 50), July 1977. (ERIC Document Reproduction Service No. ED 142 977, 33p., PC-$3.65, MF-$.91)

Anderson, R. C., & Freebody, P. Vocabulary Knowledge (No. 136), August 1979. (ERIC Document Reproduction Service No. ED 177 480, 71p., PC-$5.30, MF-$.91)

Anderson, R. C., Goetz, E. T., Pichert, J. W., & Halff, H. M. Two Faces of the Conceptual Peg Hypothesis (No. 6), January 1976. (ERIC Document Reproduction Service No. ED 134 930, 29p., PC-$3.65, MF-$.91)

Anderson, R. C., & Pichert, J. W. Recall of Previously Unrecallable Information Following a Shift in Perspective (No. 41), April 1977. (ERIC Document Reproduction Service No. ED 142 974, 37p., PC-$3.65, MF-$.91)

Anderson, R. C., Pichert, J. W., Goetz, E. T., Schallert, D. L., Stevens, K. C., & Trollip, S. R. Instantiation of General Terms (No. 10), March 1976. (ERIC Document Reproduction Service No. ED 134 933, 30p., PC-$3.65, MF-$.91)

Anderson, R. C., Pichert, J. W., & Shirey, L. L. Effects of the Reader's Schema at Different Points in Time (No. 119), April 1979. (ERIC Document Reproduction Service No. ED 169 523, 36p., PC-$3.65, MF-$.91)

Anderson, R. C., Reynolds, R. E., Schallert, D. L., & Goetz, E. T. Frameworks for Comprehending Discourse (No. 12), July 1976. (ERIC Document Reproduction Service No. ED 134 935, 33p., PC-$3.65, MF-$.91)

Anderson, R. C., Spiro, R. J., & Anderson, M. C. *Schemata as Scaffolding for the Representation of Information in Connected Discourse* (No. 24), March 1977. (ERIC Document Reproduction Service No. ED 136 236, 18p., PC-$2.00, MF-$.91)

Anderson, R. C., Stevens, K. C., Shifrin, Z., & Osborn, J. *Instantiation of Word Meanings in Children* (No. 46), May 1977. (ERIC Document Reproduction Service No. ED 142 976, 22p., PC-$2.00, MF-$.91)

Anderson, T. H. *Study Skills and Learning Strategies* (No. 104), September 1978. (ERIC Document Reproduction Service No. ED 161 000, 41p., PC-$3.65, MF-$.91)

Anderson, T. H., & Armbruster, B. B. *Studying* (No. 155), January 1980. (ERIC Document Reproduction Service No. ED 181 427, 48p., PC-$3.65, MF-$.91)

Anderson, T. H., Standiford, S. N., & Alessi, S. M. *Computer Assisted Problem Solving in an Introductory Statistics Course* (No. 56), August 1977. (ERIC Document Reproduction Service No. ED 146 563, 26p., PC-$3.65, MF-$.91)

Anderson, T. H., Wardrop, J. L., Hively, W., Muller, K. E., Anderson, R. I., Hastings, C. N., & Fredericksen, J. *Development and Trial of a Model for Developing Domain Referenced Tests of Reading Comprehension* (No. 86), May 1978. (ERIC Document Reproduction Service No. ED 157 036, 69p., PC-$5.30, MF-$.91)

Andre, M. E. D. A., & Anderson, T. H. *The Development and Evaluation of a Self-Questioning Study Technique* (No. 87), June 1978. (ERIC Document Reproduction Service No. ED 157 037, 37p., PC-$3.65, MF-$.91)

Antos, S. J. *Processing Facilitation in a Lexical Decision Task* (No. 113), January 1979. (ERIC Document Reproduction Service No. ED 165 129, 84p., PC-$6.95, MF-$.91)

Armbruster, B. B. *Learning Principles from Prose: A Cognitive Approach Based on Schema Theory* (No. 11), July 1976. (ERIC Document Reproduction Service No. ED 134 934, 48p., PC-$3.65, MF-$.91)

Armbruster, B. B., & Anderson, T. H. *The Effect of Mapping on the Free Recall of Expository Text* (No. 160), February 1980. (ERIC Document Reproduction Service No. ED 182 735, 49p., PC-$3.65, MF-$.91)

Armbruster, B. B., Stevens, R. J., & Rosenshine, B. *Analyzing Content Coverage and Emphasis: A Study of Three Curricula and Two Tests* (No. 26), March 1977. (ERIC Document Reproduction Service No. ED 136 238, 22p., PC-$2.00, MF-$.91)

Arter, J. A., & Jenkins, J. R. *Differential Diagnosis-Prescriptive Teaching: A Critical Appraisal* (No. 80), January 1978. (ERIC Document Reproduction Service No. ED 150 578, 104p., PC-$8.60, MF-$.91)

Asher, S. R. *Referential Communication* (No. 90), June 1978. (ERIC Document Reproduction Service No. ED 159 597, 71p., PC-$5.30, MF-$.91)

Asher, S. R. *Influence of Topic Interest on Black Children and White Children's Reading Comprehension* (No. 99), July 1978. (ERIC Document Reproduction Service No. ED 159 661, 35p., PC-$3.65, MF-$.91)

Asher, S. R., Hymel, S., & Wigfield, A. *Children's Comprehension of High- and Low-Interest Material and a Comparison of Two Cloze Scoring Methods* (No. 17), November 1976. (ERIC Document Reproduction Service No. ED 134 939, 32p., PC-$3.65, MF-$.91)

Asher, S. R., & Wigfield, A. *Influence of Comparison Training on Children's Referential Communication* (No. 139), August 1979. (ERIC Document Reproduction Service No. ED 177 493, 42p., PC-$3.65, MF-$.91)

Asher, S. R., & Wigfield, A. *Training Referential Communication Skills* (No. 175), July 1980. (ERIC Document Reproduction Service No. ED 191 014, 54p., PC-$5.30, MF-$.91)

Baker, L. *Processing Temporal Relationships in Simple Stories: Effects of Input Sequence* (No. 84), April 1978. (ERIC Document Reproduction Service No. ED 157 016, 54p., PC-$5.30, MF-$.91)

Baker, L. *Comprehension Monitoring: Identifying and Coping with Text Confusions* (No. 145), September 1979. (ERIC Document Reproduction Service No. ED 177 525, 62p., PC-$5.30, MF-$.91)

Baker, L., & Anderson, R. I. *Effects of Inconsistent Information on Text Processing: Evidence for Comprehension Monitoring* (No. 203), May 1981.

Baker, L., & Brown, A. L. *Metacognitive Skills and Reading* (No. 188), November 1980. (ERIC Document Reproduction Service No. ED 195 932, 74p., PC-$5.30, MF-$.91)

Baker, L., & Stein, N. L. *The Development of Prose Comprehension Skills* (No. 102), September 1978. (ERIC Document Reproduction Service No. ED 159 663, 69p., PC-$5.30, MF-$.91)

Barnitz, J. *Interrelationship of Orthography and Phonological Structure in Learning to Read* (No. 57), August 1977. (ERIC Document Reproduction Service No. ED 150 546, 62p., PC-$5.30, MF-$.91)

Barnitz, J. G. *Reading Comprehension of Pronoun-Referent Structures by Children in Grades Two, Four, and Six* (No. 117), March 1979. (ERIC Document Reproduction Service No. ED 170 731, 51p., PC-$5.30, MF-$.91)

Brewer, W. F. *Memory for the Pragmatic Implications of Sentences* (No. 65), October 1977. (ERIC Document Reproduction Service No. ED 146 564, 27p., PC-$3.65, MF-$.91)

Brewer, W. F., & Lichtenstein, E. H. *Event Schemas, Story Schemas, and Story Grammars* (No. 197), December 1980.

Brown, A. L. *Knowing When, Where, and How to Remember: A Problem of Metacognition* (No. 47), June 1977. (ERIC Document Reproduction Service No. ED 146 562, 152p., PC-$11.90, MF-$.91)

Brown, A. L. *Theories of Memory and the Problems of Development: Activity, Growth, and Knowledge* (No. 51), July 1977. (ERIC Document Reproduction Service No. ED 144 041, 59p., PC-$5.30, MF-$.91)

Brown, A. L. *Learning and Development: The Problems of Compatibility, Access, and Induction* (No. 165), March 1980. (ERIC Document Reproduction Service No. ED 184 093, 76p., PC-$6.95, MF-$.91)

Brown, A. L., & Campione, J. C. *Memory Strategies in Learning: Training Children to Study Strategically* (No. 22), March 1977. (ERIC Document Reproduction Service No. ED 136 234, 54p., PC-$5.30, MF-$.91)

Brown, A. L., & Campione, J. C. *Permissible Inferences from the Outcome of Training Studies in Cognitive Development Research* (No. 127), May 1979. (ERIC Document Reproduction Service No. ED 170 736, 34p., PC-$3.65, MF-$.91)

Brown, A. L., & Campione, J. C. *Inducing Flexible Thinking: The Problem of Access* (No. 156), January 1980. (ERIC Document Reproduction Service No. ED 181 428, 44p., PC-$3.65, MF-$.91)

Brown, A. L., Campione, J. C., & Barclay, C. R. *Training Self-Checking Routines for Estimating Test Readiness: Generalization from List Learning to Prose Recall* (No. 94), July 1978. (ERIC Document Reproduction Service No. ED 158 226, 41p., PC-$3.65, MF-$.91)

Brown, A. L., Campione, J. C., & Day, J. D. *Learning to Learn: On Training Students to Learn from Texts* (No. 189), November 1980.

Brown, A. L., & DeLoache, J. S. *Skills, Plans, and Self-Regulation* (No. 48), July 1977. (ERIC Document Reproduction Service No. ED 144 040, 66p., PC-$5.30, MF-$.91)

Brown, A. L., & French, L. A. *The Zone of Potential Development: Implications for Intelligence Testing in the Year 2000* (No. 128), May 1979. (ERIC Document Reproduction Service No. ED 170 737, 46p., PC-$3.65, MF-$.91)

Brown, A. L., & Smiley, S. S. *The Development of Strategies for Studying Prose Passages* (No. 66), October 1977. (ERIC Document Reproduction Service No. ED 145 371, 59p., PC-$5.30, MF-$.91)

Brown, A. L., Smiley, S. S., Day, J. D., Townsend, M. A. R., & Lawton, S. C. *Intrusion of a Thematic Idea in Children's Comprehension and Retention of Stories* (No. 18), December 1976. (ERIC Document Reproduction Service No. ED 136 189, 39p., PC-$3.65, MF-$.91)

Brown, A. L., Smiley, S. S., & Lawton, S. C. *The Effects of Experience on the Selection of Suitable Retrieval Cues for Studying from Prose Passages* (No. 53), July 1977. (ERIC Document Reproduction Service No. ED 144 042, 30p., PC-$3.65, MF-$.91)

Bruce, B. C. *Plans and Social Actions* (No. 34), April 1977. (ERIC Document Reproduction Service No. ED 149 328, 45p., PC-$3.65, MF-$.91)

Bruce, B. *Analysis of Interacting Plans as a Guide to the Understanding of Story Structure* (No. 130), June 1979. (ERIC Document Reproduction Service No. ED 174 951, 43p., PC-$3.65, MF-$.91)

Bruce, B. C., Collins, A., Rubin, A. D., & Gentner, D. *A Cognitive Science Approach to Writing* (No. 89), June 1978. (ERIC Document Reproduction Service No. ED 157 039, 57p., PC-$5.30, MF-$.91)

Bruce, B. C., & Newman, D. *Interacting Plans* (No. 88), June 1978. (ERIC Document Reproduction Service No. ED 157 038, 100p., PC-$6.95, MF-$.91)

Campione, J. C., Nitsch, K., Bray, N., & Brown, A. L. *Improving Memory Skills in Mentally Retarded Children: Empirical Research and Strategies for Intervention* (No. 196), December 1980.

Canney, G., & Winograd, P. *Schemata for Reading and Reading Comprehension Performance* (No. 120), April 1979. (ERIC Document Reproduction Service No. ED 169 520, 99p., PC-$6.95, MF-$.91)

Cohen, P. R., & Perrault, C. R. *Elements of a Plan-Based Theory of Speech Acts* (No. 141), September 1979. (ERIC Document Reproduction Service No. ED 177 497, 76p., PC-$6.95, MF-$.91)

Collins, A., Brown, A. L., Morgan, J. L., & Brewer, W. F. *The Analysis of Reading Tasks and Texts* (No. 43), April 1977. (ERIC Document Reproduction Service No. ED 145 404, 96p., PC-$6.95, MF-$.91)

Collins, A., Brown, J. S., & Larkin, K. M. *Inference in Text Understanding* (No. 40), December 1977. (ERIC Document Reproduction Service No. ED 150 547, 48p., PC-$3.65, MF-$.91)

Collins, A., & Smith, E. E. *Teaching the Process of Reading Comprehension* (No. 182), September 1980. (ERIC Document Reproduction Service No. ED 193 616, 43p., PC-$3.65, MF-$.91)

Davison, A. *Linguistics and the Measurement of Syntactic Complexity: The Case of Raising* (No. 173), May 1980. (ERIC Document Reproduction Service No. ED 186 848, 60p., PC-$5.30, MF-$.91)

Davison, A., Kantor, R. N., Hannah, J., Hermon, G., Lutz, R., Salzillo, R. *Limitations of Readability Formulas in Guiding Adaptations of Texts* (No. 162), March 1980. (ERIC Document Reproduction Service No. ED 184 090, 157p., PC-$11.90, MF-$.91)

Dunn, B. R., Mathews, S. R., II, & Bieger, G. *Individual Differences in the Recall of Lower-Level Textual Information* (No. 150), December 1979. (ERIC Document Reproduction Service No. ED 181 448, 37p., PC-$3.65, MF-$.91)

Durkin, D. *What Classroom Observations Reveal about Reading Comprehension Instruction* (No. 106), October 1978. (ERIC Document Reproduction Service No. ED 162 259, 94p., PC-$6.95, MF-$.91)

Fleisher, L. S., & Jenkins, J. R. *Effects of Contextualized and Decontextualized Practice Conditions on Word Recognition* (No. 54), July 1977. (ERIC Document Reproduction Service No. ED 144 043, 37p., PC-$3.65, MF-$.91)

Fleisher, L. S., Jenkins, J. R., & Pany, D. Effects on Poor Readers' Comprehension of Training in Rapid Decoding (No. 103), September 1978. (ERIC Document Reproduction Service No. ED 159 664, 39p., PC-$3.65, MF-$.91)

Freebody, P., & Anderson, R. C. Effects of Differing Proportions and Locations of Difficult Vocabulary on Text Comprehension (No. 202), May 1981.

Gearhart, M., & Hall, W. S. Internal State Words: Cultural and Situational Variation in Vocabulary Usage (No. 115), February 1979. (ERIC Document Reproduction Service No. ED 165 131, 66p., PC-$5.30, MF-$.91)

Gentner, D. On Relational Meaning: The Acquisition of Verb Meaning (No. 78), December 1977. (ERIC Document Reproduction Service No. ED 149 325, 46p., PC-$3.65, MF-$.91)

Gentner, D. Semantic Integration at the Level of Verb Meaning (No. 114), February 1979. (ERIC Document Reproduction Service No. ED 165 130, 39p., PC-$3.65, MF-$.91)

Gentner, D. Verb Semantic Structures in Memory for Sentences: Evidence for Componential Representation (No. 151), December 1979. (ERIC Document Reproduction Service No. ED 181 424, 75p., PC-$5.30, MF-$.91)

Goetz, E. T. Sentences in Lists and in Connected Discourse (No. 3), November 1975. (ERIC Document Reproduction Service No. ED 134 927, 75p., PC-$5.30, MF-$.91)

Goetz, E. T. Inferences in the Comprehension of and Memory for Text (No. 49), July 1977. (ERIC Document Reproduction Service No. ED 150 548, 97p., PC-$6.95, MF-$.91)

Goetz, E. T., Anderson, R. C., & Schallert, D. L. The Representation of Sentences in Memory (No. 144), September 1979. (ERIC Document Reproduction Service No. ED 177 527, 71p., PC-$5.30, MF-$.91)

Goetz, E. T., & Osborn, J. Procedures for Sampling Texts and Tasks in Kindergarten through Eighth Grade (No. 30), April 1977. (ERIC Document Reproduction Service No. ED 146 565, 80p., PC-$6.95, MF-$.91)

Green, G. M. Discourse Functions of Inversion Construction (No. 98), July 1978. (ERIC Document Reproduction Service No. ED 160 998, 42p., PC-$3.65, MF-$.91)

Green, G. M. Organization, Goals, and Comprehensibility in Narratives: Newswriting, a Case Study (No. 132), July 1979. (ERIC Document Reproduction Service No. ED 174 949, 66p., PC-$5.30, MF-$.91)

Green, G. M. Linguistics and the Pragmatics of Language Use: What You Know When You Know a Language . . . and What Else You Know (No. 179), August 1980. (ERIC Document Reproduction Service No. ED 193 666, 73p., PC-$5.30, MF-$.91)

Green, G. M., Kantor, R. N., Morgan, J. L., Stein, N. L., Hermon, G., Salzillo, R., & Sellner, M. B. Analysis of "Babar Loses His Crown" (No. 169), April 1980. (ERIC Document R[illegible]oduction Service No. ED 185 514, 89p., PC-$6.95, MF-$.91)

Green, G. M., Kantor, R. N., Morgan, J. L., Stein, N. L., Hermon, G., Salzillo, R., & Sellner, M. B. Analysis of "The Wonderful Desert" (No. 170), April 1980. (ERIC Document Reproduction Service No. ED 185 515, 47p., PC-$3.65, MF-$.91)

Green, G. M., Kantor, R. N., Morgan, J. L., Stein, N. L., Hermon, G., Salzillo, R., Sellner, M. B., Bruce, B. C., Gentner, D., & Webber, B. L. Problems and Techniques of Text Analysis (No. 168), April 1980. (ERIC Document Reproduction Service No. ED 185 513, 173p., PC-$11.90, MF-$.91)

Green, G. M., & Laff, M. O. Five-Year-Olds' Recognition of Authorship by Literary Style (No. 181), September 1980. (ERIC Document Reproduction Service No. ED 193 615, 44p., PC-$3.65, MF-$.91)

Grueneich, R., & Trabasso, T. The Story as Social Environment: Children's Comprehension and Evaluation of Intentions and Consequences (No. 142), September 1979. (ERIC Document Reproduction Service No. ED 177 496, 56p., PC-$5.30, MF-$.91)

Halff, H. M. Graphical Evaluation of Hierarchical Clustering Schemes (No. 1), October 1975. (ERIC Document Reproduction Service No. ED 134 926, 11p., PC-$2.00, MF-$.91)

Hall, W. S., & Dore, J. Lexical Sharing in Mother-Child Interaction (No. 161), March 1980. (ERIC Document Reproduction Service No. ED 184 066, 39p., PC-$3.65, MF-$.91)

Hall, W. S., & Guthrie, L. F. On the Dialect Question and Reading (No. 121), May 1979. (ERIC Document Reproduction Service No. ED 169 522, 32p., PC-$3.65, MF-$.91)

Hall, W. S., & Guthrie, L. F. Cultural and Situational Variation in Language Function and Use: Methods and Procedures for Research (No. 148), October 1979. (ERIC Document Reproduction Service No. ED 179 944, 49p., PC-$3.65, MF-$.91)

Hall, W. S., Linn, R. L., & Nagy, W. E. Spoken Words (No. 177), August 1980.

Hall, W. S., & Nagy, W. E. Theoretical Issues in the Investigation of Words of Internal Report (No. 146), October 1979. (ERIC Document Reproduction Service No. ED 177 526, 108p., PC-$8.60, MF-$.91)

Hall, W. S., & Tirre, W. C. The Communicative Environment of Young Children: Social Class, Ethnic, and Situational Differences (No. 125), May 1979. (ERIC Document Reproduction Service No. ED 170 788, 30p., PC-$3.65, MF-$.91)

Hansen, J., & Pearson, P. D. The Effects of Inference Training and Practice on Young Children's Comprehension (No. 166), April 1980. (ERIC Document Reproduction Service No. ED 186 839, 53p., PC-$5.30, MF-$.91)

Hayes, D. A., & Tierney, R. J. Increasing Background Knowledge through Analogy: Its Effects upon Comprehension and Learning (No. 186), October 1980. (ERIC Document Reproduction Service No. ED 195 953, 81p., PC-$6.95, MF-$.91)

Hermon, G. On the Discourse Structure of Direct Quotation (No. 143), September 1979. (ERIC Document Reproduction Service No. ED 177 495, 46p., PC-$3.65, MF-$.91)

Huggins, A. W. F. Syntactic Aspects of Reading Comprehension (No. 33), April 1977. (ERIC Document Reproduction Service No. ED 142 972, 68p., PC-$5.30, MF-$.91)

Iran-Nejad, A. The Schema: A Structural or a Functional Pattern (No. 159), February 1980. (ERIC Document Reproduction Service No. ED 181 449, 46p., PC-$3.65, MF-$.91)

Iran-Nejad, A., Ortony, A., & Rittenhouse, R. K. The Comprehension of Metaphorical Uses of English by Deaf Children (No. 184), October 1980. (ERIC Document Reproduction Service No. ED 193 618, 34p., PC-$3.65, MF-$.91)

Jenkins, J. R., & Larson, K. Evaluating Error Correction Procedures for Oral Reading (No. 55), June 1978. (ERIC Document Reproduction Service No. ED 158 224, 34p., PC-$3.65, MF-$.91)

Jenkins, J. R., & Pany, D. Curriculum Biases in Reading Achievement Tests (No. 16), November 1976. (ERIC Document Reproduction Service No. ED 134 938, 24p., PC-$2.00, MF-$.91)

Jenkins, J. R., Pany, D., & Schreck, J. Vocabulary and Reading Comprehension: Instructional Effects (No. 100), August 1978. (ERIC Document Reproduction Service No. ED 160 999, 50p., PC-$3.65, MF-$.91)

Johnston, P. Implications of Basic Research for the Assessment of Reading Comprehension (No. 206), May 1981.

Kane, J. H., & Anderson, R. C. Depth of Processing and Interference Effects in the Learning and Remembering of Sentences (No. 21), February 1977. (ERIC Document Reproduction Service No. ED 134 942, 29p., PC-$3.65, MF-$.91)

Kleiman, G. M. The Effect of Previous Context on Reading Individual Words (No. 20), February 1977. (ERIC Document Reproduction Service No. ED 134 941, 76p., PC-$6.95, MF-$.91)

Kleiman, G. M. The Prelinguistic Cognitive Basis of Children's Communicative Intentions (No. 19), February 1977. (ERIC Document Reproduction Service No. ED 134 940, 51p., PC-$5.30, MF-$.91)

Kleiman, G. M. *The Scope of Facilitation of Word Recognition from Single Word and Sentence Frame Contexts* (No. 133), July 1979. (ERIC Document Reproduction Service No. ED 174 947, 61p., PC-$5.30, MF-$.91)

Kleiman, G. M., Winograd, P. N., & Humphrey, M. M. *Prosody and Children's Parsing of Sentences* (No. 123), May 1979. (ERIC Document Reproduction Service No. ED 170 733, 28p., PC-$3.65, MF-$.91)

Linn, R. L., Levine, M. V., Hastings, C. N., & Wardrop, J. L. *An Investigation of Item Bias in a Test of Reading Comprehension* (No. 163), March 1980. (ERIC Document Reproduction Service No. ED 184 091, 97p., PC-$6.95, MF-$.91)

Mason, J. M. *Questioning the Notion of Independent Processing Stages in Reading* (No. 8), February 1976. (*Journal of Educational Psychology*, 1977, *69*, 288-297.

Mason, J. M. *Reading Readiness: A Definition and Skills Hierarchy from Preschoolers' Developing Conceptions of Print* (No. 59), September 1977. (ERIC Document Reproduction Service No. ED 145 403, 57p., PC-$5.30, MF-$.91)

Mason, J. M. *The Role of Strategy in Reading in the Mentally Retarded* (No. 58), September 1977. (ERIC Document Reproduction Service No. ED 145 406, 28p., PC-$3.65, MF-$.91)

Mason, J. M. *Prereading: A Developmental Perspective* (No. 198), February 1981.

Mason, J. M., & Au, K. H. *Learning Social Context Characteristics in Prereading Lessons* (No. 205), May 1981.

Mason, J. M., & Kendall, J. R. *Facilitating Reading Comprehension Through Text Structure Manipulation* (No. 92), June 1978. (ERIC Document Reproduction Service No. ED 157 041, 36p., PC-$3.65, MF-$.91)

Mason, J. M., Knisely, E., & Kendall, J. *Effects of Polysemous Words on Sentence Comprehension* (No. 85), May 1978. (ERIC Document Reproduction Service No. ED 157 015, 34p., PC-$3.65, MF-$.91)

Mason, J., & McCormick, C. *Testing the Development of Reading and Linguistic Awareness* (No. 126), May 1979. (ERIC Document Reproduction Service No. ED 170 735, 50p., PC-$3.65, MF-$.91)

Mason, J., Osborn, J., & Rosenshine, B. *A Consideration of Skill Hierarchy Approaches to the Teaching of Reading* (No. 42), December 1977. (ERIC Document Reproduction Service No. ED 150 549, 176p., PC-$13.55, MF-$.91)

McClure, E. *Aspects of Code-Switching in the Discourse of Bilingual Mexican-American Children* (No. 44), April 1977. (ERIC Document Reproduction Service No. ED 142 975, 38p., PC-$3.65, MF-$.91)

McClure, E., Mason, J., & Barnitz, J. *Story Structure and Age Effects on Children's Ability to Sequence Stories* (No. 122), May 1979. (ERIC Document Reproduction Service No. ED 170 732, 75p., PC-$5.30, MF-$.91)

McClure, E., & Steffensen, M. S. *A Study of the Use of Conjunctions across Grades and Ethnic Groups* (No. 158), January 1980. (ERIC Document Reproduction Service No. ED 182 688, 43p., PC-$3.65, MF-$.91)

McConkie, G. W. *Evaluating and Reporting Data Quality in Eye Movement Research* (No. 193), December 1980.

McConkie, G. W., Hogaboam, T. W., Wolverton, G. S., Zola, D., & Lucas, P. A. *Toward the Use of Eye Movements in the Study of Language Processing* (No. 134), August 1979. (ERIC Document Reproduction Service No. ED 174 968, 48p., PC-$3.65, MF-$.91)

McConkie, G. W., & Zola, D. *Language Constraints and the Functional Stimulus in Reading* (No. 194), December 1980.

Morgan, J. L. *Two Types of Convention in Indirect Speech Acts* (No. 52), July 1977. (ERIC Document Reproduction Service No. ED 145 405, 40p., PC-$3.65, MF-$.91)

Nash-Webber, B. *Anaphora: A Cross-Disciplinary Survey* (No. 31), April 1977. (ERIC Document Reproduction Service No. ED 144 039, 43p., PC-$3.65, MF-$.91)

Nash-Webber, B. L. *Inferences in an Approach to Discourse Anaphora* (No. 77), January 1978. (ERIC Document Reproduction Service No. ED 150 552, 30p., PC-$3.65, MF-$.91)

Nash-Webber, B., & Reiter, R. *Anaphora and Logical Form: On Formal Meaning Representation for Natural Language* (No. 36), April 1977. (ERIC Document Reproduction Service No. ED 142 973, 42p., PC-$3.65, MF-$.91)

Nezworski, T., Stein, N. L., & Trabasso, T. *Story Structure Versus Content Effects on Children's Recall and Evaluative Inferences* (No. 129), June 1979. (ERIC Document Reproduction Service No. ED 172 187, 49p., PC-$3.65, MF-$.91)

Nicholson, T., Pearson, P. D., & Dykstra, R. *Effects of Embedded Anomalies and Oral Reading Errors on Children's Understanding of Stories* (No. 118), March 1979. (ERIC Document Reproduction Service No. ED 169 524, 43p., PC-$3.65, MF-$.91)

Nolan, S. D., Tanenhaus, M. K., & Seidenberg, M. S. *Multiple Code Activation in Word Recognition: Evidence from Rhyme Monitoring* (No. 204), May 1981.

Ortony, A. *Names, Descriptions, and Pragmatics* (No. 7), February 1976. (ERIC Document Reproduction Service No. ED 134 931, 25p., PC-$2.00, MF-$.91)

Ortony, A. *Remembering and Understanding Jabberwocky and Small-Talk* (No. 28), March 1977. (ERIC Document Reproduction Service No. ED 137 753, 36p., PC-$3.65, MF-$.91)

Ortony, A. *Beyond Literal Similarity* (No. 105), October 1978. (ERIC Document Reproduction Service No. ED 166 635, 58p., PC-$5.30, MF-$.91)

Ortony, A. *Some Psycholinguistic Aspects of Metaphor* (No. 112), January 1979. (ERIC Document Reproduction Service No. ED 165 115, 38p., PC-$3.65, MF-$.91)

Ortony, A. *Understanding Metaphors* (No. 154), January 1980. (ERIC Document Reproduction Service No. ED 181 426, 52p., PC-$5.30, MF-$.91)

Ortony, A., Reynolds, R. E., & Arter, J. A. *Metaphor: Theoretical and Empirical Research* (No. 27), March 1977. (ERIC Document Reproduction Service No. ED 137 752, 63p., PC-$5.30, MF-$.91)

Ortony, A., Schallert, D. L., Reynolds, R. E., & Antos, S. J. *Interpreting Metaphors and Idioms: Some Effects of Context on Comprehension* (No. 93), July 1978. (ERIC Document Reproduction Service No. ED 157 042, 41p., PC-$3.65, MF-$.91)

Pany, D., & Jenkins, J. R. *Learning Word Meanings: A Comparison of Instructional Procedures and Effects on Measures of Reading Comprehension with Learning Disabled Students* (No. 25), March 1977. (ERIC Document Reproduction Service No. ED 136 237, 34p., PC-$3.65, MF-$.91)

Pearson, P. D., Hansen, J., & Gordon, C. *The Effect of Background Knowledge on Young Children's Comprehension of Explicit and Implicit Information* (No. 116), March 1979. (ERIC Document Reproduction Service No. ED 169 521, 26p., PC-$3.65, MF-$.91)

Pearson, P. D., Raphael, T., TePaske, N., & Hyser, C. *The Function of Metaphor in Children's Recall of Expository Passages* (No. 131), July 1979. (ERIC Document Reproduction Service No. ED 174 950, 41p., PC-$3.65, MF-$.91)

Pichert, J. W. *Sensitivity to What is Important in Prose* (No. 149), November 1979. (ERIC Document Reproduction Service No. ED 179 946, 64p., PC-$5.30, MF-$.91)

Pichert, J. W., & Anderson, R. C. *Taking Different Perspectives on a Story* (No. 14), November 1976. (ERIC Document Reproduction Service No. ED 134 936, 30p., PC-$3.65, MF-$.91)

Raphael, T. E., Myers, A. C., Freebody, P., Tirre, W. C., & Fritz, M. *Contrasting the Effects of Some Text Variables on Comprehension and Ratings of Comprehensibility* (No. 190), December 1980.

Reder, L. M. *Comprehension and Retention of Prose: A Literature Review* (No. 108), November 1978. (ERIC Document Reproduction Service No. ED 165 114, 116p., PC-$8.60, MF-$.91)

Reichman, R. *Conversational Coherency* (No. 95), July 1978. (ERIC Document Reproduction Service No. ED 159 658, 86p., PC-$6.95, MF-$.91)

Reynolds, R. E., & Anderson, R. C. *Influence of Questions on the Allocation of Attention during Reading* (No. 183), October 1980. (ERIC Document Reproduction Service No. ED 193 617, 44p., PC-$3.65, MF-$.91)

Reynolds, R. E., & Ortony, A. Some Issues in the Measurement of Children's Comprehension of Metaphorical Language (No. 172), May 1980. (ERIC Document Reproduction Service No. ED 185 542, 42p., PC-$3.65, MF-$.91)

Reynolds, R. E., Standiford, S. N., & Anderson, R. C. Distribution of Reading Time When Questions are Asked about a Restricted Category of Text Information (No. 83), April 1978. (ERIC Document Reproduction Service No. ED 153 206, 34p., PC-$3.65, MF-$.91)

Reynolds, R. E., Taylor, M. A., Steffensen, M. S., Shirey, L. L., & Anderson, R. C. Cultural Schemata and Reading Comprehension (No. 201), April 1981.

Royer, J. M. Theories of Learning Transfer (No. 79), January 1978. (ERIC Document Reproduction Service No. ED 149 326, 55p., PC-$5.30, MF-$.91)

Royer, J. M., & Cunningham, D. J. On the Theory and Measurement of Reading Comprehension (No. 91), June 1978. (ERIC Document Reproduction Service No. ED 157 040, 63p., PC-$5.30, MF-$.91)

Royer, J. M., Hastings, C. N., & Hook, C. A Sentence Verification Technique for Measuring Reading Comprehension (No. 137), August 1979. (ERIC Document Reproduction Service No. ED 176 234, 34p., PC-$3.65, MF-$.91)

Rubin, A. D. A Theoretical Taxonomy of the Differences between Oral and Written Language (No. 35), January 1978. (ERIC Document Reproduction Service No. ED 150 550, 61p., PC-$5.30, MF-$.91)

Rubin, A. D., Bruce, B. C., & Brown, J. S. A Process-Oriented Language for Describing Aspects of Reading Comprehension (No. 13), November 1976. (ERIC Document Reproduction Service No. ED 136 188, 41p., PC-$3.65, MF-$.91)

Schallert, D. L. Improving Memory for Prose: The Relationship between Depth of Processing and Context (No. 5), November 1975. (ERIC Document Reproduction Service No. ED 134 929, 37p., PC-$3.65, MF-$.91)

Schallert, D. L., Kleiman, G. M., & Rubin, A. D. Analyses of Differences between Written and Oral Language (No. 29), April 1977. (ERIC Document Reproduction Service No. ED 144 038, 33p., PC-$3.65, MF-$.91)

Schwartz, R. M. Strategic Processes in Beginning Reading (No. 15), November 1976. (ERIC Document Reproduction Service No. ED 134 937, 19p., PC-$2.00, MF-$.91)

Schwartz, R. M. Relation of Context Utilization and Orthographic Automaticity in Word Identification (No. 45), May 1977. (ERIC Document Reproduction Service No. ED 137 762, 27p., PC-$3.65, MF-$.91)

Schwartz, R. M. Levels of Processing: The Strategic Demands of Reading Comprehension (No. 135), August 1979. (ERIC Document Reproduction Service No. ED 177 471, 45p., PC-$3.65, MF-$.91)

Seidenberg, M. S., Tanenhaus, M. K., & Leiman, J. M. The Time Course of Lexical Ambiguity Resolution in Context (No. 164), March 1980. (ERIC Document Reproduction Service No. ED 184 092, 58p., PC-$5.30, MF-$.91)

Shoben, E. J. *Choosing a Model of Sentence Picture Comparisons: A Reply to Catlin and Jones* (No. 81), February 1978. (ERIC Document Reproduction Service No. ED 150 577, 30p., PC-$3.65, MF-$.91)

Shoben, E. J., Rips, L. J., & Smith, E. E. *Issues in Semantic Memory: A Response to Glass and Holyoak* (No. 101), August 1978. (ERIC Document Reproduction Service No. ED 159 662, 85p., PC-$6.95, MF-$.91)

Siegel, M. A. *Teacher Behaviors and Curriculum Packages: Implications for Research and Teacher Education* (No. 9), April 1976. (ERIC Document Reproduction Service No. ED 134 932, 42p., PC-$3.65, MF-$.91)

Smiley, S. S., Oakley, D. D., Worthen, D., Campione, J. C., & Brown, A. L. *Recall of Thematically Relevant Material by Adolescent Good and Poor Readers as a Function of Written Versus Oral Presentation* (No. 23), March 1977. (ERIC Document Reproduction Service No. ED 136 235, 23p., PC-$2.00, MF-$.91)

Smith, E. E. *Organization of Factual Knowledge* (No. 185), October 1980. (ERIC Document Reproduction Service No. ED 195 954, 109p., PC-$8.60, MF-$.91)

Spiro, R. J. *Inferential Reconstruction in Memory for Connected Discourse* (No. 2), October 1975. (ERIC Document Reproduction Service No. ED 136 187, 81p., PC-$6.95, MF-$.91)

Spiro, R. J. *Etiology of Reading Comprehension Style* (No. 124), May 1979. (ERIC Document Reproduction Service No. ED 170 734, 21p., PC-$2.00, MF-$.91)

Spiro, R. J. *Prior Knowledge and Story Processing: Integration, Selection, and Variation* (No. 138), August 1979. (ERIC Document Reproduction Service No. ED 176 235, 41p., PC-3.32, MF-$.91)

Spiro, R. J. *Schema Theory and Reading Comprehension: New Directions* (No. 191), December 1980.

Spiro, R. J., & Esposito, J. J. *Superficial Processing of Explicit Inferences in Text* (No. 60), December 1977. (ERIC Document Reproduction Service No. ED 150 545, 27p., PC-$3.65, MF-$.91)

Spiro, R. J., & Taylor, B. M. *On Investigating Children's Transition from Narrative to Expository Discourse: The Multidimensional Nature of Psychological Text Classification* (No. 195), December 1980.

Spiro, R. J., & Tirre, W. C. *Individual Differences in Schema Utilization During Discourse Processing* (No. 111), January 1979. (ERIC Document Reproduction Service No. ED 166 651, 29p., PC-$3.65, MF-$.91)

Steffensen, M. S. *Bereiter and Engelmann Reconsidered: The Evidence from Children Acquiring Black English Vernacular* (No. 82), March 1978. (ERIC Document Reproduction Service No. ED 153 204, 31p., PC-$3.65, MF-$.91)

Steffensen, M. S., & Guthrie, L. F. *Effect of Situation on the Verbalization of Black Inner-City Children* (No. 180), September 1980. (ERIC Document Reproduction Service No. ED 193 614, 37p., PC-$3.65, MF-$.91)

Steffensen, M. S., Jogdeo, C., & Anderson, R. C. *A Cross-Cultural Perspective on Reading Comprehension* (No. 97), July 1978. (ERIC Document Reproduction Service No. ED 159 660, 41p., PC-$3.65, MF-$.91)

Steffensen, M. S., Reynolds, R. E., McClure, E., & Guthrie, L. F. *Black English Vernacular and Reading Comprehension: A Cloze Study of Third, Sixth, and Ninth Graders* (No. 199), February 1981.

Stein, N. L. *How Children Understand Stories: A Developmental Analysis* (No. 69), March 1978. (ERIC Document Reproduction Service No. ED 153 205, 68p., PC-$5.30, MF-$.91)

Stein, N. L., & Goldman, S. *Children's Knowledge about Social Situations: From Causes to Consequences* (No. 147), October 1979. (ERIC Document Reproduction Service No. ED 177 524, 54p., PC-$5.30, MF-$.91)

Stein, N. L., & Nezworski, T. *The Effects of Organization and Instructional Set on Story Memory* (No. 68), January 1978. (ERIC Document Reproduction Service No. ED 149 327, 41p., PC-$3.65, MF-$.91)

Stein, N. L., & Trabasso, T. *What's in a Story: An Approach to Comprehension and Instruction* (No. 200), April 1981.

Straker, D. Y. *Situational Variables in Language Use* (No. 167), April 1980. (ERIC Document Reproduction Service No. ED 185 619, 49p., PC-$3.65, MF-$.91)

Tanenhaus, M. K., Flanigan, H., & Seidenberg, M. S. *Orthographic and Phonological Activation in Auditory and Visual Word Recognition* (No. 178), August 1980. (ERIC Document Reproduction Service No. ED 193 620, 46p., PC-$3.65, MF-$.91)

Tanenhaus, M. K., & Seidenberg, M. S. *Discourse Context and Sentence Perception* (No. 176), July 1980. (ERIC Document Reproduction Service No. ED 191 015, 45p., PC-$3.65, MF-$.91)

Thieman, T. J., & Brown, A. L. *The Effects of Semantic and Formal Similarity on Recognition Memory for Sentences in Children* (No. 76), November 1977. (ERIC Document Reproduction Service No. ED 150 551, 26p., PC-$3.65, MF-$.91)

Tierney, R. J., & Cunningham, J. W. *Research on Teaching Reading Comprehension* (No. 187), November 1980. (ERIC Document Reproduction Service No. ED 195 946, 125p., PC-$8.60, MF-$.91)

Tierney, R. J., & Mosenthal, J. *Discourse Comprehension and Production: Analyzing Text Structure and Cohesion* (No. 152), January 1980. (ERIC Document Reproduction Service No. ED 179 945, 84p., PC-$6.95, MF-$.91)

Tirre, W. C., Freebody, P., & Kaufman, K. *Achievement Outcomes of Two Reading Programs: An Instance of Aptitude-Treatment Interaction* (No. 174), June 1980. (ERIC Document Reproduction Service No. ED 193 619, 34p., PC-$3.65, MF-$.91)

Tirre, W. C., Manelis, L., & Leicht, K. L. *The Effects of Imaginal and Verbal Strategies on Prose Comprehension in Adults* (No. 110), December 1978. (ERIC Document Reproduction Service No. ED 165 116, 27p., PC-$3.65, MF-$.91)

Trabasso, T. *On the Making of Inferences During Reading and Their Assessment* (No. 157), January 1980. (ERIC Document Reproduction Service No. ED 181 429, 38p., PC-$3.65, MF-$.91)

Wardrop, J. L., Anderson, T. H., Hively, W., Anderson, R. I., Hastings, C. N., & Muller, K. E. *A Framework for Analyzing Reading Test Characteristics* (No. 109), December 1978. (ERIC Document Reproduction Service No. ED 165 117, 65p., PC-$5.30, MF-$.91)

Wigfield, A., & Asher, S. R. *Age Differences in Children's Referential Communication Performance: An Investigation of Task Effects* (No. 96), July 1978. (ERIC Document Reproduction Service No. ED 159 659, 31p., PC-$3.65, MF-$.91)

Winograd, P., & Johnston, P. *Comprehension Monitoring and the Error Detection Paradigm* (No. 153), January 1980. (ERIC Document Reproduction Service No. ED 181 425, 57p., PC-$5.30, MF-$.91)

Woods, W. A. *Multiple Theory Formation in High-Level Perception* (No. 38), April 1977. (ERIC Document Reproduction Service No. ED 144 020, 58p., PC-$5.30, MF-$.91)

Zehler, A. M., & Brewer, W. F. *Acquisition of the Article System in English* (No. 171), May 1980. (ERIC Document Reproduction Service No. ED 186 907, 51p., PC-$5.30, MF-$.91)